Volume 41 June 1955 Number 6

Proceedings of the
National Academy of Sciences
of the United States of America

The Proceedings of the National Academy of Sciences

OFFICERS OF THE ACADEMY

DETLEV W. BRONK
President

GEORGE W. CORNER
Vice President

JOHN KIRKWOOD
Foreign Secretary

ALEXANDER WETMORE
Home Secretary

WILLIAM J. ROBBINS
Treasurer

EDITORIAL BOARD OF THE PROCEEDINGS

LINUS PAULING
Chairman

ALEXANDER WETMORE, *Home Secretary*
EDWIN B. WILSON, *Managing Editor*
JOHN KIRKWOOD, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS	G. W. BEADLE	GREGORY BREIT
R. E. CLELAND	E. A. DOISY	L. C. DUNN
W. W. RUBEY	G. C. EVANS	W. F. LIBBY
HARLOW SHAPLEY	C. H. GRAHAM	J. VON NEUMANN
F. E. TERMAN	J. T. PATTERSON	PAUL WEISS

Subscription price is $12.50 per volume (12 issues); the price of a single issue is $1.25. Make all remittances payable to The University of Chicago Press in U.S. currency or its equivalent. Authorized agent for the British Commonwealth, except North America and Australasia: Cambridge University Press, Bentley House, 200 Euston Road, London, N. W. 1, England.

Editorial correspondence should be addressed to the PROCEEDINGS of the NATIONAL ACADEMY OF SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Business correspondence should be addressed to the Journals Division, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Subscribers are requested to notify The University of Chicago Press and their local postmaster immediately of change of address.

Microfilms of complete volumes of this journal are available to regular subscribers only and may be obtained at the end of the volume year from University Microfilms, 313 N. First Street, Ann Arbor, Michigan.

Entered as second-class matter June 21, 1920 at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized June 18, 1920.

PRINTED IN THE U. S. A.

THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
is published for
THE NATIONAL ACADEMY OF SCIENCES
by
THE UNIVERSITY OF CHICAGO PRESS

Proceedings of the
NATIONAL ACADEMY OF SCIENCES

Volume 41 · Number 6 · June 15, 1955

*ZINC, A COMPONENT OF YEAST ALCOHOL DEHYDROGENASE**

By Bert L. Vallee and Frederic L. Hoch

BIOPHYSICS RESEARCH LABORATORY OF THE DEPARTMENT OF MEDICINE, HARVARD MEDICAL SCHOOL, AND PETER BENT BRIGHAM HOSPITAL, BOSTON, MASSACHUSETTS

Communicated by F. O. Schmitt, March 21, 1955

The ADH of yeast, first crystallized by Negelein and Wulff in 1937,[1] depends upon the presence of DPN for activity.[2] The mechanisms of its enzymatic actions have been studied.[3, 4]

The present work extends previous observations concerning the chemical composition,[1] physical-chemical interactions, and enzymological behavior of yeast ADH. The apoenzyme has been shown to contain uniformly large amounts of zinc firmly bound to the protein. The metal is a functional component of the molecule in its enzymatic activity.

Methods—Enzyme activity of ADH was determined by measuring the rate of formation or disappearance of DPNH.[5] A Cary recording spectrophotometer was employed to record the change in optical density at 340 mμ. The slope of the initial linear part of the reaction was measured and expressed as activity units V: change in absorbance ($\Delta\epsilon$) at 340 mμ per minute. These values are presented as specific activity units V per milligram of ADH. For the measurement of the rate of formation of DPNH, the following 3.0-ml. reaction mixture was employed at 23° C.: DPN, 5 μmoles; ethanol, 1,000 μmoles; 1.0 ml. of 0.1 M pyrophosphate buffer, pH 8.8; approximately 5μgm. of ADH in 0.2 ml. The 3.0-ml. mixture for the reaction DPNH → DPN at 23° C. contained approximately 3 μmoles DPNH;[6] 25 μmoles acetaldehyde; 1.0 ml. of 0.1 M phosphate buffer, pH 6.5; ADH as above. Inhibition of ADH activity was expressed as percentage of the uninhibited control reaction: $(V_i/V_0) \times 100$.

Zinc: Two dithizone methods differing in the mode of sample preparation were employed.[7]

Emission spectrography: Samples were dry-ashed in a thermostatically controlled electric muffle furnace. Internal standards were added to the ash. Aliquots were sparked in porous cup electrodes. Jarrell-Ash "Varisources" were employed to generate and control the spark. The two spectrographs were 21′ Wadsworth mountings but differed in the characteristics of their gratings. One instrument was provided with a 15,000-line-per-inch grating reflecting maximally in the first-order green, while the other 15,000-line-per-inch grating reflected maximally in the first-order violet. The reciprocal linear dispersions were 5.18 and 5.25 A/mm, respectively. Eastman Kodak 103-O photographic plates were em-

ployed throughout. Lines and backgrounds were measured densitometrically. Working curves were prepared with spectroscopically pure chemicals. Zinc was determined both chemically and spectrographically to insure accuracy.

All analytical procedures were carried out with precautions against metal contamination. Water was obtained by slow passage through a mixed ion exchange resin bed containing IR 120 and IRA 420 (Rohm and Haas). The effluent had a specific resistance of at least 1.5×10^6 ohms. Reagents were freed of metals when necessary and stored in acid-cleaned polyethylene bottles throughout.

Results.—Table 1 shows analyses on eight different crystalline preparations of highly purified yeast ADH. Preparations Nos. 6 and 7 were dialyzed against large volumes of metal-free distilled water for 24 hours at 0° C.; No. 8 was recrystallized four times.

TABLE 1

METAL CONTENT OF DIFFERENT CRYSTALLINE YEAST ADH PREPARATIONS
(All Values in Micrograms of Metal per Gram of Protein)

Metal	Exp. 8-225 (1)	Exp. 26-156 (2)	Exp. 8-200 (3)	Exp. 3-178 (4)	Exp. 23-91 (5)	Exp. 23–135 (6) (Dialyzed)	Exp. 23–148 (7) (Dialyzed)	Exp. 26–165 (8) (Recrystallized)
Zinc	1,440	1,600	1,660	1,910	1,800	2,050	1,900	1,800
Magnesium	296	130	1,180	95	630	28	25	2.4
Calcium	105	0*	39	<1	679	<1	<1	0
Aluminum	48	500	79	53	34	0	0	0
Barium	20	28	11	27	0	0	0	9.1
Strontium	2	0	4	0	0	0	0	0
Lead	0	0	45	0	0	0	0	0
Cadmium	0	0	13	0	0	0	0	0
Chromium	0	0	8	0	0	0	0	0
Iron	81	0	80	0	0	0	0	0
Copper	†	†	165	†	†	†	†	†
Moles zinc/mole protein	3.3	3.7	3.8	4.4	4.1	4.7	4.4	4.1
Moles magnesium/mole protein	1.8	0.80	7.3	0.59	3.9	0.17	0.15	0.015

* 0 = Not detectable. † Not done.

Zinc, the major metallic constituent, varies in concentration from 1,440 to 2,050 μgm. of zinc per gram of protein. The molar ratios of zinc to protein, shown in Table 1, line 12, range from 3.3 to 4.7, averaging 4.1 moles of zinc per mole of ADH, based on a molecular weight of 150,000.[3] The zinc content of one crystalline preparation was lowered from 2,210 to 1,720 μgm. of zinc per gram of protein by dialysis against OP, corresponding to 3.9 moles per mole of protein. The concentrations of magnesium, the only other element consistently present, vary widely in different preparations.

Table 2 shows data on metal content and activity obtained on consecutive fractions during a purification of ADH from yeast.[3, 5] Specific activity (Table 2, col. A) rises from 17.2×10^6 in the materials initially extracted from yeast (line 1) to 550×10^6 *V* per gram of protein in the fourth crystals (line 11). Concomitantly, zinc content (col. C) rises from 362 (line 1) to 1,800 (line 11) μgm. per gram of protein.

The activity:zinc ratio (col. B) reflects the relationship of the activity to the metal, rising from 47.5×10^3 in the extract (line 1) to 305×10^3 *V* per microgram of zinc in the fourth crystals (line 11).

Magnesium, the element present in largest concentration in the extract (line 1), falls to 2.4 μgm. per gram of protein in the fourth crystals (line 11), a 10,000-fold decrease. Activity and magnesium content vary inversely in the course of purification (cols. A and D). All other elements are also either reduced in concentration or remain at their initially insignificant levels (cols. E–H). The second crystals are contaminated with aluminum and barium (line 10).

Figure 1 shows the effects of the preincubation of 5×10^{-3} *M* OP with ADH for one hour at 0° C. at the indicated pH levels. Below pH 7.5, the enzyme is completely inhibited by OP and less so with rising pH of the incubation medium. Figure 1 also shows spectrophotometric data on the effect of pH on the formation of Zn-OP complexes for comparison.[8]

TABLE 2

ACTIVITY AND METAL CONTENT OF FRACTIONS IN THE COURSE OF PURIFICATION OF ADH FROM YEAST

	Fraction	A Specific Activity (V/Gm P + 10^6)	B Activity: Zinc (V/μgm × 10^3)	C Zinc (μgm/ Gm P)	D Magnesium (μgm/ Gm P)	E Calcium (μgm/ Gm P)	F Barium (μgm/ Gm P)	G Aluminum (μgm/ Gm P)	H Iron (μgm) Gm P/
1.	Extraction + 55° precipitation; supernatant	17.2	47.5	362	29,000	2,800	10	63	530
2.	Acetone precipitation; supernatant 1	14.0	24.1	582	2,200	0*	39	17	200
3.	† Acetone precipitation; supernatant 2	0	0	7,695	2,100	0	0	0	0
4.	Dialysis supernatant	24.8	91.5	271	4,100	0	10	12	86
5.	† $(NH_4)_2SO_4$ precipitation; supernatant 1	1.5	6.6	227	0	150	2.2	5.3	130
6.	$(NH_4)_2SO_4$ precipitation; precipitate 1	110	199	554	161	15	0.9	1.1	24
7.	$(NH_4)_2SO_4$ precipitation; supernatant 2	74.0	128	580	620	33	1.7	1.0	32
8.	† Crystals 1; supernatant	44.1	126	351	370	0	3.5	3.0	0
9.	Crystals 1	242	266	910	270	0	0	0	0
10.	Crystals 2	555	290	1,910	95	<1	27	53	0
11.	Crystals 4‡	550	305	1,800	2.4	0	9.1	0	0

Present: K, Na, P; absent: Ag, B, Cd, Co, Cr, Li, Mn, Mo, Ni, Pb, Sr, Ti.
* 0 = Not detectable.
† Discard.
‡ The fourth crystals were more than 85 per cent pure with respect to protein. We are indebted to Dr. Hans Neurath, Department of Biochemistry, University of Washington, for ultracentrifuge and electrophoresis analyses.

The effects of various concentrations of OP, 8-OHQ, 8-OHQ5SA, αα-D, DZ, and TU on the oxidation of ethanol are shown in Figure 2; the conditions of preincubation are indicated.

Figure 2*a* shows similar data on the reduction of acetaldehyde. The enzyme was preincubated with varying concentrations of OP at pH 6.5. The concentration curve is very similar in shape and close in absolute values to that obtained for the effect of OP on the oxidation of ethanol (see Fig. 2).

With the agents studied, inhibition does not occur without preincubation (V at 23° C.). Figure 3 shows ADH activity as a function of time in the presence and absence of 2×10^{-3} *M* OP at 0° C., pH 7.5. In five hours, activity of the uninhibited enzyme falls slowly to 80 per cent of the original value, while in the presence of OP it falls to 17.5 per cent during the oxidation of ethanol (DPN → DPNH). During the same period, activity of the uninhibited enzyme falls to 72.5 per cent,

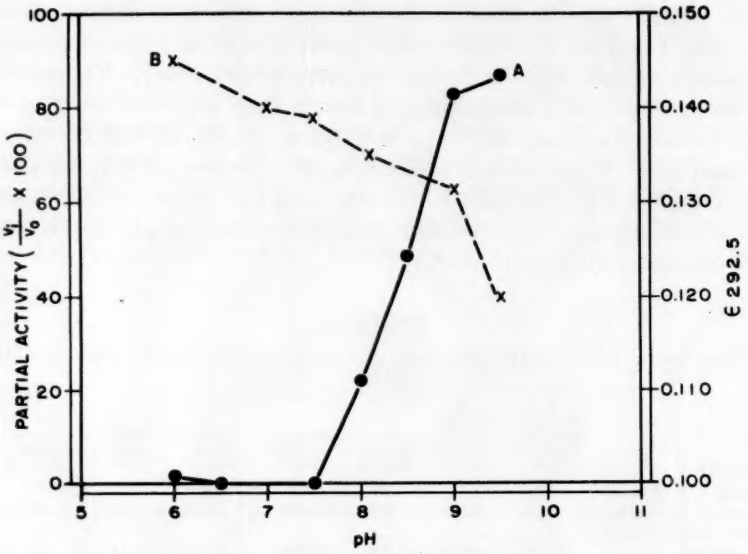


FIG. 1.—*A*, left ordinate: Effect of pH of preincubation with OP on inhibition of ADH activity. Preincubation conditions: ADH ± 5×10^{-3} *M* OP, 0° C., 60 minutes. Activity measurements: The rate of DPN → DPNH, pH 8.8, 23° C., is measured. Each point represents an activity measurement after preincubation with the inhibitors (V_i) as a percentage of the uninhibited control (V_0).

B, right ordinate: Effect of pH on the formation of $[Zn\ (OP)_n]^{++}$ complexes. The absorption at 292.5 mμ measures the formation of $[Zn\ (OP)_n]^{++}$ complexes; points are calculated from the data of McClure and Banks, U.S. Atomic Energy Commission Bull. ISC-164, 1951.

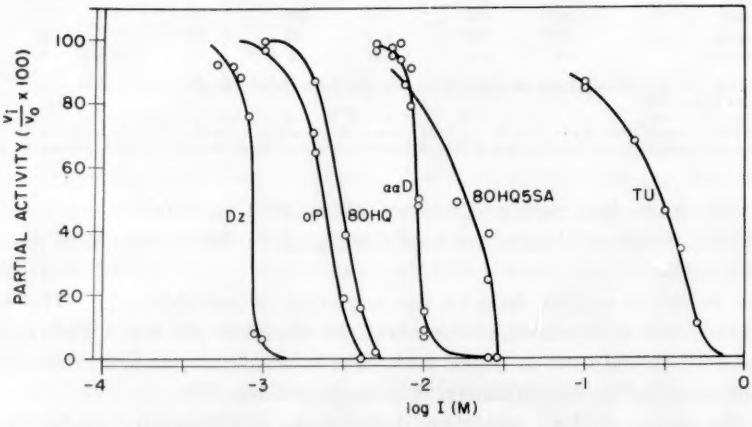


FIG. 2.—Effect of concentration of inhibitors on ADH activity, following preincubation. Preincubation conditions: ADH ± inhibitor, pH 7.5, 0° C., 60 minutes. Activity measurements: The rate of DPN → DPNH, pH 8.8, 23° C., is measured. The partial activity after inhibition, $(V_i/V_0) \times 100$, is plotted as a function of the logarithm of the molar inhibitor concentration ($\log I$).

and in the presence of OP to 18.2 per cent, of the original value during the reduction of acetaldehyde (DPNH → DPN). In the presence of OP the activity of ADH falls much more rapidly during the first few hours than without it. Thereafter, the rates of inactivation of control and inhibited enzyme proceed nearly parallel. Sixty minutes at 0° C. was chosen as a standard condition for studies of inhibition of the enzyme.

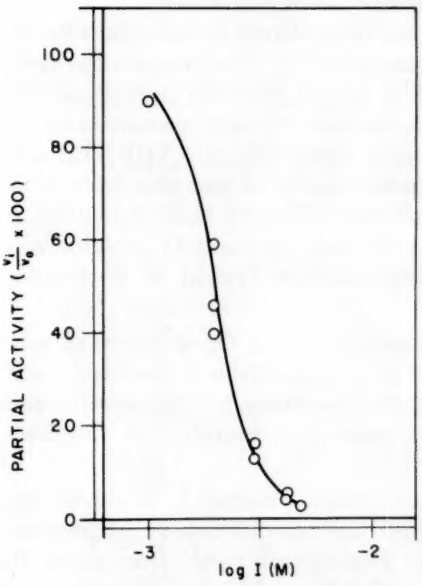


FIG. 2a.—Effect of concentration of OP on ADH activity. Preincubation conditions: ADH ± OP, pH 6.5, 60 minutes. Activity measurements: The rate of DPNH → DPN, pH 6.5, 23° C., is measured.

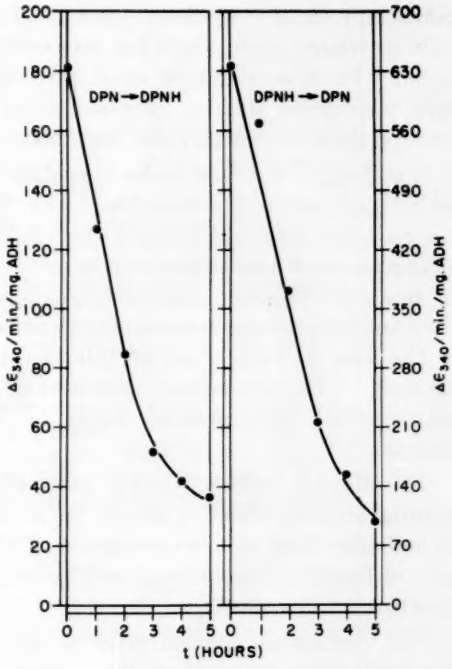


FIG. 3.—Effect of time of preincubation on inhibition of ADH activity by OP. Preincubation conditions: ADH ± 2×10^{-3} *M* OP, pH 7.5, 0° C. Activity measurements: Aliquots are removed from the same preincubation mixture at the times indicated. The rate of DPN → DPNH at pH 8.8, 23° C., or the rate of DPNH → DPN at pH 6.5, 23° C., is measured.

The higher the temperature, the greater the rate of inactivation of both the uninhibited and inhibited enzyme (Fig. 4). The rates of inactivation at 37° and 25° C. are very much higher than at 0° C.

When 2×10^{-3} *M* Zn^{++} (always as $ZnCl_2$) is added at the time at which the OP-inhibited enzyme still exhibits 18 per cent of initial activity, an immediate return to 59 per cent of the uninhibited enzymatic activity is observed (Fig. 5); 5×10^{-3} *M* Zn^{++} does not reverse the inhibition further. The addition of 2×10^{-3} *M* Zn^{++} to ADH produces a decrease to 26 per cent of control activity in 18 minutes.

Inactivation can be *prevented* completely when the ferroine group of OP is occupied by Zn^{++} or Cu^{++} ions before the enzyme is exposed to the inhibitor. Addi-

tion to the enzyme of 5×10^{-3} *M* OP exposed for 5 minutes to the same concentration of Zn^{++} results in 96 per cent of normal activity, while 5×10^{-3} *M* OP inhibits ADH to 30 per cent of normal activity.

Discussion.—The average molar ratio of zinc to protein in ADH is 4.1. It appears that four atoms of zinc are present in one molecule of ADH apoenzyme. The zinc in these crystalline preparations is firmly bound: the zinc-protein bond is maintained against competitive physical-chemical factors involved in fractionation, potentially capable of dissociating it. Recrystallization or dialysis against water fails to remove zinc. In some preparations more than 0.2 per cent of zinc was found. In these instances small amounts of zinc can be removed by relatively mild procedures, such as dialysis against OP. Such dialyses have not, thus far, lowered the zinc content below four moles of zinc per mole of protein.

It is thought that four moles of zinc are bound firmly to the protein, constituting an integral part of the molecule in the "natural state".[9] This complex has been assigned the empirical formula $[(ADH)Zn_4]$, in accord with the theoretical considerations and nomenclature to be presented elsewhere.[10] Zinc, present in excess of this concentration, represents contamination extraneous to $[(ADH)Zn_4]$ and does not increase enzymatic activity, nor does the removal of such zinc decrease it.

The data in Table 2 extend the conclusions reached on the basis of analysis of crystals. The concomitant increases of specific activity (activity:protein) and zinc content (zinc:protein) imply an interdependence typical of a metalloenzyme.[11, 12]

The discard fractions during purification contain zinc. The presence of substantial concentrations of zinc in the absence of activity (Table 2, fractions 3 and 5) indicates that zinc extraneous to $[(ADH)Zn_4]$ is removed. The activity and zinc of fraction 8 probably reflect incomplete removal of crystals from this fraction by centrifugation.

The magnesium concentration of the yeast extract (fraction 1) is almost one hundred times as high as the corresponding zinc concentration. Magnesium concentrations in the same fraction of other preparations varied from about 10 to 30 mgm. per gram of protein, giving a basis for the variable concentrations encountered in crystals. During purification the magnesium concentration decreases 10,000-fold, while specific activity increases about 30-fold. Magnesium, calcium, aluminum, barium, and iron are neither structurally nor functionally associated with ADH.

Preincubation of ADH with certain chelating agents produces marked inhibition of enzymatic activity. Since data on the physical chemistry of zinc 1,10-phenanthrolinates, $[Zn(OP)_n]^{++}$, are available,[8, 13, 14] OP was studied most intensively. The association constants, obtained by spectrophotometry, are shown in Table 3. Similar data have been obtained by potentiometric titrations (unpublished experiments).

OP has been employed for the analytical determination of zinc in the absence of iron, the only metal with a known association constant for this ligand higher than that of zinc. The absence of iron from purified ADH (Tables 1 and 2) and the inhibition of ADH activity of OP support the analytical data. Addition of $[Zn(OP)_n]^{++}$ or $[Cu(OP)_n]^{++}$ to the enzyme does not inhibit it, indicating that the

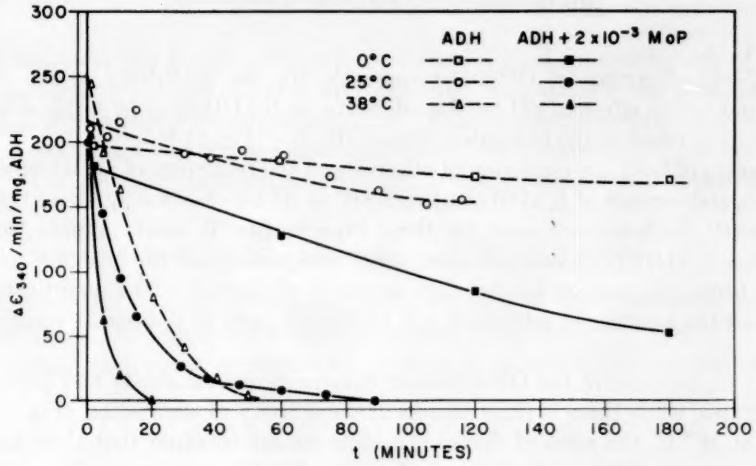


FIG. 4.—Effect of temperature and time of preincubation on inhibition of ADH activity by OP. All open symbols and broken lines represent rates of the control reaction in the absence of OP; all solid symbols and solid lines represent rates of the reaction in the presence of OP. Different temperatures are represented by the shape of symbols as indicated. Preincubation conditions: ADH ± 2 × 10^{-3} *M* OP, pH 7.5. Activity measurements: Aliquots are removed from each preincubation mixture at the times indicated, and the rate of DPN → DPNH, pH 8.8, 23° C., is determined.

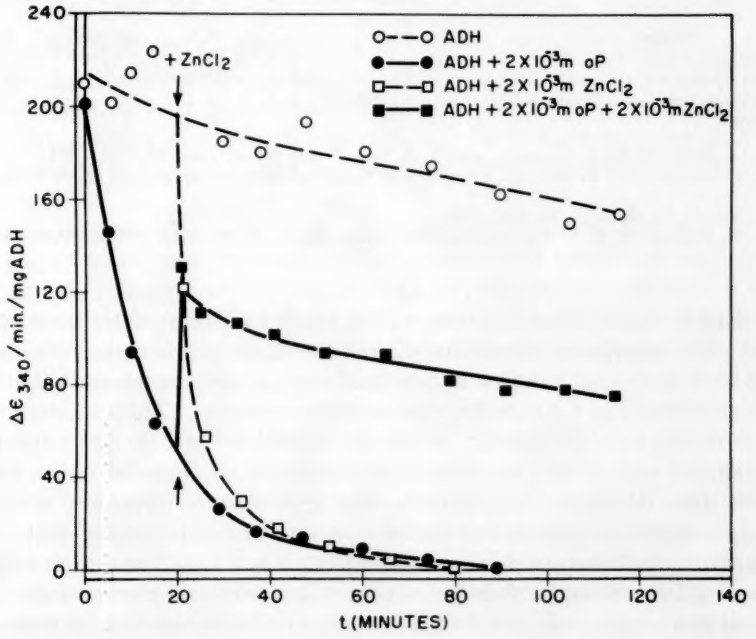


FIG. 5.—Reversal of inhibition of ADH activity by OP on addition of Zn^{++} (as $ZnCl_2$). All broken lines represent rates of the control reaction in the absence of OP, all solid lines represent rates of the reaction in the presence of OP. The addition of Zn^{++} is marked by arrows, and rates in the presence of Zn^{++} are shown as square symbols. Preincubation conditions: ADH ± 2 × 10^{-3} *M* OP, pH 7.5, 25° C. Activity measurements: Aliquots are removed from each preincubation mixture at the times indicated, and the rate of DPN → DPNH, pH 8.8, 23° C., is ascertained.

—N=C—C=N-group of OP[15] is responsible for the inhibitory action of this compound. The effect of pH on the inhibition of [(ADH)Zn_4] by OP is similar in manner to its effect on the formation of $[Zn(OP)_n]^{++}$ (Fig. 1).[8]

The rates of both the oxidation of ethanol and the reduction of acetaldehyde fall on prolonged contact of [(ADH)Zn_4] with OP at 0° C. Since aliquots of the same enzyme-OP mixture were used for these experiments, it would appear that OP acts upon [(ADH)Zn_4] and not upon other components of the reaction mixture. This is further supported by the data shown in Figure 2*a*. The quantitative aspects and the kinetics of inhibition will be the subjects of a separate communication.

At 0° C., activity of the OP-inhibited enzyme decreases slowly and linearly for 180 minutes, while there is little change in the activity of the control (Fig. 4). At 25° or at 38° C. the rates of decrease in activity are so rapid that their accurate measurement is difficult, and, particularly at 38° C., any comparison with the rapidly deteriorating controls becomes difficult. Accurate measurements could be made conveniently with a one-hour preincubation period at 0° C. The concentration of agents required to produce 50 per cent inhibition is in the order DZ < OP < 8-OHQ < $\alpha\alpha$-D < 8-OHQ5SA < TU, under these conditions.

TABLE 3

Reagent	Symbol Used in Text	pk_1	pk_1k_2	$pk_1k_2k_3$	Reference
1,10-Phenanthroline	OP	6.4	12.2	17.0	1
8-Hydroxyquinoline	8-OHQ	10.9	20.8	...	2
8-Hydroxyquinoline-5-sulfonic acid	8-OHQ5SA	8.4	15.1	...	3
Dithizone	DZ	..	20.1	...	4

1. I. M. Kolthoff, D. L. Leussing, and T. S. Lee, *J. Am. Chem. Soc.*, **73**, 390, 1951.
2. A. E. Martell and M. Calvin, *Chemistry of the Metal Chelate Compounds* (New York, 1952), p. 557.
3. A. Albert, *Biochem. J.*, **54**, 646, 1953.
4. A. K. Babko and A. T. Philipenko, *Zhur. Anal. Khim.*, **2**, 33, 1947; *Chem. Abstr.*, **43**, 5344, 1949.

The degree of inhibition may be related to the stability of the zinc complex formed. The association constants of some of these compounds with Zn^{++} in Table 3 have been used as a first approximation to gauge enzyme inhibition and to relate it to stability of the expected zinc-inhibitor complex. The inhibitory effects, while consistent with the stability of the zinc complexes, are no direct function of this parameter and present the difficulties commonly encountered in the interpretation of data on mixed complexes.[16] The geometric arrangement of the zinc atoms with respect to protein and ligand molecules, the steric and chemical factors contributed by the reactive, polar groups of protein and ligand, and their respective charges would significantly alter the constants derived in simpler systems. These analytical and enzymological data lead to the conclusion that zinc is a structural and functional component of ADH and participates in the mechanism of its enzymatic action. The four metal atoms are firmly bound to the protein apoenzyme. Pending the results of investigations in progress, this discussion will assume that the four zinc atoms are bound to the protein in an equivalent manner and that each atom of zinc acts independently of the other three in the catalytic action of alcohol dehydrogenase.

One molecule of yeast ADH has been shown to bind four molecules of DPN.[3] Potentiometric titrations of DPN and $Zn(NO_3)_2$ demonstrate the formation of a Zn++(DPN) complex (unpublished data). It is thought, therefore, that each zinc atom serves as a locus of reversible attachment for a molecule of the coenzyme, DPN. The general interaction is formulated as follows, under certain experimental conditions:

$$[(ADH)Zn_4] + 4(DPN) \rightleftarrows [(ADH)Zn_4](DPN)_4. \quad (1)$$

$[(ADH)Zn_4](DPN)_4$ is the active complex composed of the apoenzyme, zinc, and the coenzyme, the brackets denoting structural association.[10]

Enzymatic activity is decreased by incubation of the metalloenzyme with an inhibitor, I:

$$[(ADH)Zn_4] + 4I \rightleftarrows [(ADH)Zn_4]I_4. \quad (2)$$

n moles of I are bound through covalent and/or ionic bonds to each zinc atom. For the purpose of the presentation of the reactions, n has been set equal to 1. This reaction should result in an inactive, reversibly dissociable ZnI complex. Combination of equations (1) and (2) depicts the complete reaction in the presence of an inhibitor:

$$2[(ADH)Zn_4] + 4(DPN) + 4I \rightleftarrows [(ADH)Zn_4](DPN)_4 + [(ADH)Zn_4]I_4. \quad (3)$$

The existence of the two equilibria should lead to competitive inhibition between DPN and I. Preliminary work has, indeed, shown that the inhibition of ADH by OP is competitive with DPN (unpublished data), lending support to the formulation of the reaction as shown in equations (1)–(3). Apparently, chelate ligands exert their effects and are bound to $[(ADH)Zn_4]$ in a manner similar to that employed for DPN.

These considerations imply that the association between the zinc moiety of ADH and an inhibitor is reversible. This hypothesis is based on the thought that the zinc-protein bond is much stronger than any formed between zinc and an inhibitor. If it were postulated that the zinc-protein bond is broken by a ligand of high affinity for zinc, the reaction should result in the removal of zinc from the protein, and activity would be lost:

$$[(ADH)Zn_4] + 4I \overset{?}{\rightleftarrows} (ADH) + 4ZnI. \quad (4)$$

Activity could only be restored by addition of Zn^{++} to (ADH) to reconstitute the original compound. Experience with most other metalloenzymes[10, 12] makes this appear improbable. The mechanism of inhibition, therefore, depends upon the equilibrium constants of equations (2) and (4) for any given inhibitor.

The inhibition of alcohol dehydrogenase activity by OP has been reversible under experimental conditions employed thus far. With OP as the inhibitor, equation (2) apparently prevails. In view of the high association constants of the $[Zn(OP)_n]^{++}$ complexes,[13] the corresponding constants of the $[(ADH)Zn_4]$ complex must be very high indeed.

Since zinc is bound so firmly to the protein but the groups to which it is bound have not been verified, one can only speculate at this time concerning their identity.

It has been postulated that sulfhydryl groups bind DPN to liver ADH.[17] The existence of some 22 free –SH groups in yeast ADH has been reported.[18] The zinc atom itself may be bound to ADH through a thiol bond. The present findings do not rule out the existence of two linkages between ADH and DPN: one is definitely through a zinc atom of $[(ADH)Zn_4]$.

The experiment described in Figure 5 was designed to test further the formulations of the reactions given in equations (1)–(4). Zinc and OP form complexes in solution:

$$Zn^{++} + n(OP) \rightleftarrows [Zn(OP)_n]^{++} \qquad (5)$$

Such complexes should form when Zn^{++} is added to $[(ADH)Zn_4](OP)_4$ (eq. [2]), competing for OP bound to the metalloenzyme. The competition of the equilibrium shown in equation (5) with that of equation (2) should result in restoration of activity:

$$[(ADH)Zn_4](OP)_4 + 4Zn^{++} \rightleftarrows [(ADH)Zn_4] + 4[Zn(OP)_n]^{++} \qquad (6)$$

Figure 5 demonstrates the operation of this mechanism. The addition of Zn^{++} rapidly restores and maintains activity at 59 per cent of the control level. Enzymatic activity is not restored completely, however, under the conditions of the experiment. The following possibilities have been considered to explain the incomplete restoration of activity: (*a*) The inactive metalloenzyme-inhibitor complex may be dissociated incompletely (eq. [6]). (*b*) Zinc may be removed from the apoenzyme (eq. [4]). (*c*) OP may first remove some zinc from the metalloenzyme. Zinc ions, thereafter, only partially reconstitute the $[(ADH)Zn_4]$ complex. (*d*) Zinc ions may combine with DPN, making it unavailable. (*e*) Zinc ions may interact with enzymatically reactive end groups of the apoenzyme. The irreversible removal of zinc (possibility *b*) cannot be ruled out, even though experiments specifically designed to remove zinc from the apoenzyme have not been successful thus far (possibilities *b* and *c*). For this reason the reactivation of zinc-free (ADH) by Zn^{++} has not lent itself to experimental verification. The incomplete dissociation of $[(ADH)Zn_4](OP)_4$ (possibility *a*) is at present hypothesized to be the cause for the incomplete restoration of activity. Since zinc ions themselves inhibit the enzymatic reaction (Fig. 5), the incomplete restoration of activity may be accounted for in part by the mechanisms proposed under *d* and *e*, which are believed to be responsible for the effect of Zn^{++} on $[(ADH)Zn_4]$.

Zinc ions and complexes do not absorb light in the visible region of the spectrum and therefore do not draw attention to themselves, as do iron and copper, for instance. The absence of changes in its valence coupled with the stability of its complexes may be a clue to the function of zinc in this and similar systems.

It has been reported that *Neurospora crassa*, grown on zinc-deficient media, was devoid of ADH activity,[19] not restored by the addition of Zn^{++}. The *structural* association of zinc with the ADH apoenzyme was not considered as a possible explanation for the resultant functional derangements,[19] which were attributed to "an indirect influence on the synthesis of the apoenzyme."[20] In the absence of zinc, the present data imply that this organism is unable to form the functional metalloenzyme molecule $[(ADH)Zn_4]$. Apparently, the ADH of *Neurospora*, like the ADH of yeast, is a zinc metalloenzyme.

The molecular weight of ADH of horse liver has been reported to be 73,000.[21] Two DPN molecules per molecule of protein are bound.[21] The characteristics of the two enzymes differ in many respects.[22] One can only speculate as to their cause. Solely on the basis of molar proportions of protein to cofactor, the presence of two moles of zinc per mole of protein might be expected. The known high zinc content of liver,[23] of which ADH has been stated to be 1 per cent of the total protein weight,[24] lends support to this assumption.

It has been reported[25] and confirmed[26] that retina contains high concentrations of zinc. Horse liver ADH oxidizes vitamin A_1 and reduces retinene, probably being identical with retinene reductase.[27] If mammalian ADH is found to be a zinc enzyme, enzymological and analytical observations on liver and retina, and those on yeast ADH here reported, may be shown to have an analogous structural denominator.

Summary.—The ADH of yeast is a zinc metalloenzyme containing four moles of zinc firmly bound to one mole of protein. The activity of the enzyme is directly dependent on zinc. Enzymatic action is inhibited by dithizone, 1,10-phenanthroline, 8-hydroxyquinoline, 8-hydroxyquinoline-5-sulfonic acid, α,α'-dipyridyl, and thiourea. The inhibition is reversible and probably is competitive with DPN. The mechanisms of the enzymatic reaction, inactivation, and reversal of inhibition have been formulated in terms of the existence of a structural metalloenzyme entity. This has been assigned the empirical formula $[(ADH)Zn_4]$ in conformity with a general scheme of notation of metalloenzyme structure. The active metalloenzyme-coenzyme complex is represented by $[(ADH)Zn_4](DPN)_4$.

We are greatly indebted to Mrs. Alice Abrahamian, Mr. Thomas L. Coombs, Mrs. Anna Kakatsakis, Miss Flora Lerner, and Miss Bessie Zotos for excellent technical assistance during the course of this work, and to the Worthington Biochemical Corporation for their generous co-operation.

* A preliminary report has been published: B. L. Vallee and E. L. Hoch, *J. Am. Chem. Soc.* **77**, 821, 1955. These studies were supported by grants-in-aid from the Office of Naval Research, Contract No. NR 119-277; the National Institutes of Health; the Research Corporation, New York; the Howard Hughes Medical Institute; and the Rockefeller Foundation.

The following abbreviations will be used: ADH, alcohol dehydrogenase; DPN, diphosphopyridine nucleotide; DPNH, reduced diphosphopyridine nucleotide; OP, 1,10-phenanthroline; 8-OHQ, 8-hydroxyquinoline; $\alpha\alpha$-D, α,α'-dipyridyl; 8-OHQ5SA, 8-hydroxyquinoline-5-sulfonic acid; TU, thiourea.

[1] E. Negelein and H.-J. Wulff, *Biochem. Z.*, **293**, 351, 1937.

[2] B. Andersson, *Z. physiol. Chem.*, **225**, 57, 1934.

[3] J. E. Hayes, Jr., and S. F. Velick, *J. Biol. Chem.*, **207**, 225, 1954.

[4] F. H. Westheimer, H. F. Fisher, E. E. Conn, and B. Vennesland, *J. Am. Chem. Soc.*, **73**, 2403, 1951; M. E. Pullman, A. San Pietro, and S. P. Colowick, *J. Biol. Chem.*, **206**, 129, 1954.

[5] E. Racker, *J. Biol. Chem.*, **184**, 313, 1950.

[6] R. Bonnichsen, *Acta Chem. Scand.*, **4**, 714, 1950.

[7] B. L. Vallee and J. G. Gibson II, *J. Biol. Chem.*, **176**, 435, 1948; F. L. Hoch and B. L. Vallee, *J. Biol. Chem.*, **181**, 295, 1949.

[8] J. H. McClure and C. V. Banks, U.S. Atomic Energy Commission Bull. ISC-164, 1951.

[9] J. T. Edsall (ed.), *Enzymes and Enzyme Systems: Their State in Nature* (Cambridge, Mass., 1951).

[10] B. L. Vallee (in preparation for publication).

[11] B. L. Vallee, *Sci. Monthly*, **62**, 368, 1951; "Trace Elements and Pathologic States," T. R. Harrison (ed.), *Principles of Internal Medicine*, (2d. ed.; Philadelphia, 1954), p. 437.
[12] B. L. Vallee, F. L. Hoch, and W. L. Hughes, Jr., *Arch. Biochem. and Biophys.*, **48**, 347, 1954.
[13] I. M. Kolthoff, D. L. Leussing, and T. S. Lee, *J. Am. Chem. Soc.*, **73**, 390, 1951.
[14] W. W. Brandt, F. P. Dwyer, and E. C. Gyarfas, *Chem. Rev.*, **54**, 959, 1954.
[15] G. F. Smith, *Anal. Chem.*, **26**, 1934, 1954.
[16] I. M. Klotz and W.-C. Loh Ming, *J. Am. Chem. Soc.*, **76**, 805, 1954.
[17] H. Theorell and R. Bonnichsen, *Acta Chem. Scand.*, **5**, 329, 1951.
[18] E. S. G. Barron and S. Levine, *Arch. Biochem. and Biophys.*, **41**, 175, 1952.
[19] A. Nason, N. O. Kaplan, and S. P. Colowick, *J. Biol. Chem.*, **188**, 397, 1951.
[20] W. D. McElroy and A. Nason, *Ann. Rev. Plant Physiol.*, **5**, 1, 1954.
[21] H. Theorell and R. Bonnichsen, *Acta Chem. Scand.*, **5**, 1105, 1951.
[22] R. Bonnichsen, *4. Colloq. Ges. physiol. Chem.*, p. 151, April, 1953.
[23] M. Leiner and G. Leiner, *Naturwissenschaften*, **29**, 763, 1941.
[24] R. Bonnichsen, *Acta Chem. Scand.*, **4**, 715, 1950.
[25] M. Leiner and G. Leiner, *Biol. Zentr.*, **64**, 293, 1944.
[26] J. M. Bowness, R. A. Morton, M. H. Shakir, and A. L. Stubbs, *Biochem. J.*, **51**, 521, 1952.
[27] A. F. Bliss, *Biol. Bull.*, **97**, 221, 1949.

PRODUCTION OF ABNORMAL (PSYCHOTIC?) BEHAVIOR IN MICE WITH LYSERGIC ACID DIETHYLAMIDE, AND ITS PARTIAL PREVENTION WITH CHOLINERGIC DRUGS AND SEROTONIN

BY D. W. WOOLLEY*

ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH, NEW YORK

Communicated April 7, 1955

The production of hallucinations in men by the feeding of ergot alkaloids, and especially of the derivative lysergic acid diethylamide (LSD-25), has been known for many years.[1] The hallucinogenic action of LSD-25, particularly, has been much studied recently because such minute doses (1 μgm. per kilogram) are effective and because the mental aberration called forth by this compound is said to resemble schizophrenia. Our demonstration in 1953[2, 3] and that of Gaddum[4] that the ergot alkaloids are naturally occurring antimetabolites of the hormone serotonin, capable of antagonizing the action of this hormone on smooth muscle preparations, suggested that these alkaloids, and especially LSD-25, acted on the brain to antagonize the normal functioning of serotonin just as they did on muscles.[5, 6] The deficiency of serotonin thus induced in the brain by the drug was considered as a possible cause of the hallucinations, and it was suggested that some of the naturally occurring mental diseases might similarly arise from a lack of serotonin in the brain. This lack would result not from drug action but rather from a failure to form enough serotonin in this organ.

In order to test such a hypothesis, it was necessary to induce a recognizable mental disorder with LSD-25 and to overcome it with serotonin. Because it was already quite clear that serotonin administered peripherally did not enter the brain readily,[5] various means of direct introduction into the central nervous system seemed a prerequisite of the experiment, and for this reason, as well as for others, a test system involving animals rather than men was essential. Intracerebral injections into men for experimental purposes would be out of the question.

The difficulties of working with, or even recognizing, a mental disorder in a laboratory animal must be plain to all. The main hope would be to cause a clearly defined and readily recognizable change in behavior by administration of LSD-25. If one were then fortunate enough to be able to recognize this change as a part of the animal's normal response to a specific emergency not induced by drugs, one might make a reasonable deduction that an hallucination had been induced by the drug. Of course, one could never be sure, because hallucinations are necessarily subjective phenomena, but at least one could make a reasonable start.

When a mouse is placed at one edge of a large glass plate, and this edge is raised (not too rapidly) so that the plate forms an inclined plane of steadily increasing pitch with the mouse facing downhill, an angle of inclination of the plane is soon approached at which the animal will slide. The mouse does not attempt to turn around, but, rather, it attempts to brace itself so that it will not move. It looks apprehensively and rapidly from side to side but apparently does not judge it wise to essay escape by trying to turn around and walk uphill. It spreads its forelegs, and, especially, it spreads its toes on the forepaws in an obvious effort to prevent slippage. It pushes back gingerly with the forelegs. As the inclination of the plate is increased, the mouse slides down it, still exhibiting the behavior just described.

Effects of LSD-25 on Mice.—This behavior could be induced in a mouse maintained on a level surface merely by injection of LSD-25. In all these experiments young adult females of the Swiss albino strain of the kind raised in this Institute were used. They weighed approximately 25 gm. each. The LSD-25 was dissolved in Ringer's solution and injected intraperitoneally.[7] The mice were placed in individual 2-liter glass beakers on a level surface.

Approximately 3 minutes after the injection of the LSD-25, the mouse began to look rapidly and apprehensively from side to side. Soon the toes were spread wide, and the animal began to push backward with its forelegs. Then it started to walk backward. It thus moved backward, either by pushing or by actually walking until its rear end encountered the wall of the beaker. When one placed a finger at the mouse's tail and tried to push it forward, considerable force was felt as the animal tried to push backward. This pushing and walking backward continued, sometimes intermittently, for 10–30 minutes. Often the hind legs were spread far apart, and always the mouse held itself close to the floor. The head continued to move from side to side. With sufficient LSD-25 the walking backward was always seen. With smaller doses the pushing backward occurred, but not the actual walking. In this latter situation a mouse would sometimes pivot slowly around one hind leg. In the normal course of events, the agitation of the mouse increased, and marked tremors ensued. These usually appeared about 7–10 minutes after the injection. Very soon after the injection, the ears became markedly red, and, as this sign subsided, the hair stood on end. These last three signs—tremors, red ears, and horripilation—were not seen in normal mice placed on the inclined plane as described earlier in this paper. They thus constituted additional manifestations not seen in normal mice faced with the prospect of sliding downhill. Furthermore, these three signs proved to be the most difficult ones to eradicate with serotonin and cholinergic drugs. For the present study, the exhibition of agitation, and especially the pushing and walking backward, were taken as the chief criteria that some sort of hallucination was being experienced by the animal.

The quantity of LSD-25 required to produce this type of behavior can be judged from Table 1. It is important to note that 100 per cent of the animals were thus affected by doses subsequently used in this study (100 μgm., 4 mg. per kilogram). Consequently, no statistical problem arose in evaluation of the effects of antagonistic drugs. It was judged important to arrange the test in this fashion, so that reasonable numbers of animals could be observed. This dosage was much greater than that employed for production of hallucinations in men, but the frank signs in the mice were far greater than those observable in men.

TABLE 1

DOSAGE OF LSD-25 REQUIRED TO CAUSE MICE TO WALK BACKWARD

Intraperitoneal LSD-25 (μgm. per Mouse)	No. of Animals	No. Walking Backward	Intraperitoneal LSD-25 (μgm. per Mouse)	No. of Animals	No. Walking Backward
300	3	1	50	8	7
150	40	40	25	4	3
100	19	19	5	2	0

Table 1 also will show that, if the dose was too large, the walking backward was not always seen. Mice so treated were more prostrate than those getting smaller amounts and gave the impression of not being sufficiently co-ordinated to be able to walk.

Prevention of the Walking Backward with Serotonin plus Cholinergic Drugs.—The intracerebral injection of serotonin plus carbamylcholine, or of serotonin plus physostigmine, prevented the walking and pushing backward of LSD-25-treated mice in about one-quarter of the cases. For such an experiment a young adult female mouse which had been caged solitarily for 24 hours was anesthetized with ether, and the solution of serotonin plus cholinergic drug (in Ringer's solution) was injected with a 27-gauge needle in such a way that most of it was delivered into the lateral ventricle of the brain. The volume of solution so introduced never exceeded 0.03 cc. A second mouse received an equal volume of plain Ringer's solution. When the animals had regained consciousness (about 1 minute), they were each given intraperitoneally 100 μgm. of LSD-25. This was enough to cause 100 per cent of the controls to walk backward. They were then placed in glass beakers and observed. The data from many such experiments are summarized in Table 2. It was found best to carry out each experiment with not more than three animals and then to repeat it many times. If larger numbers of animals were under observation at one time, some of the signs escaped attention. The usual procedure was to have three mice, the third one receiving serotonin plus LSD-25 without cholinergic drug, or cholinergic drug plus LSD-25 without serotonin. Suitable controls without the LSD-25 also were set up frequently.

The data of Table 2 show that the walking backward of the LSD-25 mice was prevented in about one-quarter of the cases by intracerebral injection of 20 μgm. of serotonin plus 0.2 μgm. of carbamylcholine. This latter compound is a derivative of acetylcholine, which, because it is resistant to cholinesterase, persists longer than does acetylcholine itself. In fact, when acetylcholine plus serotonin was injected, no protection against the effects of LSD-25 was found.

The data of Table 2 also show that the carbamylcholine could be replaced by physostigmine. As is well known, this drug inhibits cholinesterase, thereby allowing the accumulation of endogenous acetylcholine.

Intracerebral serotonin alone did not prevent the manifestations caused by LSD-25. Injections of as much as 30 μgm. of it intracerebrally failed to prevent the walking backward. The cholinergic drug (carbamylcholine or physostigmine) by itself in the amounts used above did not protect against the effects of LSD-25. However, if larger amounts of these drugs were used (e.g., 0.6 μgm. of carbamylcholine or 3 μgm. of physostigmine), then some protection was afforded by them alone. With these large doses, the mice exhibited the signs of poisoning with such drugs, i.e., excessive salivation and tonic spasm of the tail muscles.

It was very important to keep the mice caged solitarily for at least a day before the experiments with intracerebral drugs plus LSD-25. Before this fact was discovered, great variation was found in the amount of cholinergic drug required. If mice were taken from a number of colony cages and then placed in a box together while the experiment was being made ready, it was usually found that 0.3 μgm. of carbamylcholine, and sometimes as little as 0.05 μgm., was sufficient by itself, without serotonin, to prevent the walking backward. Under such circum-

TABLE 2

ANTAGONISM BETWEEN LSD-25 AND SEROTONIN PLUS CHOLINERGIC DRUGS IN THE ABNORMAL BEHAVIOR OF MICE

(All Mice Given 100 μgm. of LSD-25 Each Intraperitoneally)

I.C. Serotonin (μgm.)	I.C. Carbamylcholine (μgm.)	I C. Physostigmine (μgm.)	No. of Mice	No. Walking Backward	No. Pushing but Not Walking Backward
0	0	0	19	19	0
20	0	0	5	5	0
0	0.3	0	3	3	0
0	0.2	0	13	13	0
30	0.3	0	11	8	1
20	0.2	0	23	14	3
20	0.1	0	2	0	2
20	0.05	0	5	3	2
5	0.2	0	5	4	1
1	0.2	0	2	2	0
20	0	1.0	5	2	3
0	0	1.0	5	4	1

stances, the demonstration of the effect of serotonin was rather difficult, because one never knew just how much cholinergic drug would be required so that the mixture of it with serotonin would protect while the drug alone would not. The introduction of solitary caging for at least a day overcame this problem. Apparently, the excitement engendered by suddenly creating a new social group may have influenced the endogenous production of acetylcholine. The minimal amounts of serotonin and of cholinergic drugs needed to prevent the walking backward were determined by varying the one while the other was held constant. The data of Table 2 will indicate that the minimal amounts were 20 μgm. of serotonin and 0.1 μgm. of carbamylcholine. Significant protection was not obtained with less than 0.3 μgm. of physostigmine plus the serotonin.

Many efforts were made to find why all the mice were not protected by this mixture of drugs. There was some evidence that the place of injection into the brain decided whether or not protection would be achieved. However, no certain proof of this was established. As far as could be determined by gross examination, the intracerebral injections were delivered into the lateral ventricle. It may have

been that some particular site was required for the injection to be effective and that this was reached only in those mice which did show protection.

Failure to Protect Mice against LSD-25 by the Drugs Administered Peripherally.—Many experiments were performed in an attempt to prevent the LSD-25 signs by peripheral administration of serotonin (up to 100 μgm.) plus carbamylcholine or physostigmine. These uniformly failed. Even when the cholinergic drugs were given intracerebrally and the serotonin intraperitoneally, there was no evidence of protection. These failures added to the previous evidence[5] which indicated that peripherally injected serotonin did not enter the brain, at least in sufficient amounts to be demonstrable.

Effects on Other Signs Resulting from LSD-25 Treatment.—Although the walking and pushing backward of mice caused by LSD-25 were prevented in some cases by intracerebral serotonin plus cholinergic drugs, the animals so protected did not behave as normal mice. The reddening of the ears was diminished but not abolished. The agitation was largely controlled, but the tremors were not completely abolished. The most difficult sign to overcome was the marked raising of the hair. The mixture of serotonin and cholinergic drug did not prevent this. Furthermore, the mice protected from the walking and pushing backward were less active than normal individuals. They would walk forward if touched, and very occasionally of their own accord, but for the most part they just sat still.

Failure to Protect Mice against LSD-25 with Nembutal.—Because mice which had received serotonin plus carbamylcholine or physostigmine intracerebrally were more lethargic than those which received only saline, the objection might be raised that they were partially anesthetized and consequently incapable of walking backward. In order to examine this possibility, mice were partially anesthetized with nembutal. Some were given just enough to bring about full anesthesia and were used just as they were regaining consciousness. Others were given just enough nembutal to make them very lethargic but not yet anesthetized. In these conditions they were given 100 μgm. of LSD-25. Without exception, they became more active than they had been previously, and walked backward in characteristic fashion. This showed that just plain partial anesthesia was not enough to protect against the abnormal behavior elicited by LSD-25.

Serotonin versus LSD-25 on Artery Rings.—Although data have been published which show the antagonism between serotonin and some of the ergot alkaloids as measured on segments of carotid arteries[2, 3] and although Gaddum *et al.*[4] have given evidence of a highly specific antagonism of the same sort between serotonin and LSD-25 in a few smooth muscle preparations, no data have yet been printed on the antagonism between these last two substances as measured on arteries. Therefore, the following data are recorded. The experiments with segments of sheep carotid arteries were conducted according to the directions of Woolley and Shaw.[8] The results are shown in Table 3. They indicate that this ergot alkaloid derivative is much more active as an antiserotonin than is ergotamine or ergotoxin.[3] Since LSD-25 showed an inhibition index of 1, it was among the most potent antiserotonins thus far described.

TABLE 3

SEROTONIN VERSUS LSD-25 IN THE CONTRACTION OF CAROTID ARTERY RINGS

Serotonin (μgm./cc.)	LSD (μgm./cc.)	Decrease in Diameter of Ring (Per Cent)	Serotonin (μgm./cc.)	LSD (μgm./cc.)	Decrease in Diameter of Ring (Per Cent)
0	0	0	0.2	0.3	12
0.2	0	31	0.2	1.0	6
0.2	0.03	23	0	10.0	−3
0.2	0.1	15			

DISCUSSION

The reason why both serotonin and cholinergic drugs were required to prevent the abnormal (psychotic?) behavior induced in mice by LSD-25 was not clear. In isolated smooth muscles, serotonin alone was sufficient. One possible explanation is to be found in the work of Greig and Carter.[9] They showed that cholinergic drugs increase the penetration of substances into slices of brain. It may well be that cholinergic drugs increase the penetrability of brain cells to serotonin and that without them in the present experiments the serotonin did not adequately reach its site of action. This could be true even though the serotinin was injected into the lateral ventricle of the brain. In the experiments of Greig and Carter the brain slices were actually in contact with the substance, the penetration of which was increased by acetylcholine. Thus, while it is clear from the results of the present study that serotonin does not readily pass from the peripheral circulation into the brain (cf. also Woolley and Shaw[5]), the problem of getting it into the cells even after it has entered the fluid spaces of the central nervous system may still be an important one.

Another bit of evidence which may bear on the relationship of acetylcholine to serotonin is the cholinergic effects of serotonin on isolated guinea pig ileum described by Roche e Silva and Ribiero do Valle.[10] There would seem to be some interrelationship between these two hormones in muscles and presumably also in nerve tissue, but the nature of this relationship is not yet clear.

Is the abnormal behavior of the LSD-25 mice the result of a pyschosis? One can never tell with certainty what a mouse is thinking of, and consequently it is not possible to answer this question objectively. However, because of the resemblance of the behavior of an LSD-25 mouse to that of a normal mouse on an inclined plane, the use of the working hypothesis seems to have some basis for justification. Nevertheless, one must recognize that the LSD-25 may be affecting muscles and motor nerves without any influence on mental processes. It is only the requirement for an assay method using experimental animals rather than human beings that encourages one to consider such a test.

The working hypothesis which the results of the present study suggest is that LSD-25 causes a mental disturbance in mice by induction of a serotonin deficiency in the central nervous system. This disturbance can be partially overcome in some animals by injection of serotonin plus carbamylcholine or physostigmine. These compounds must be given directly into the brain, because they do not seem to pass the blood-brain barrier. It is hoped that it will be possible to discover a relative of serotonin which, while still acting like the hormone, is able to pass this

barrier. One would thus achieve a substance which could be administered in an unobjectionable way (peripherally) and which would be expected to control the mental aberration.

After this work had been completed, a paper appeared[11] describing effects of LSD-25 on Siamese fighting fish. It was interesting to note that LSD-25 caused these animals to swim backward especially by action of the pectoral fins.

SUMMARY

The administration of lysergic acid diethylamide (LSD-25) to mice caused them to behave in a way similar to untreated mice faced with the prospect of sliding down an inclined plane. It was believed that the drug induced in the mice an hallucination of sliding down an inclined plane. This abnormal behavior was prevented in some of the animals by injection of serotonin plus carbamylcholine into the lateral ventricle of the brain. The carbamylcholine could be replaced by the cholinesterase inhibitor physostigmine. The serotonin plus carbamylcholine had to be given intracerebrally. Intraperitoneal injection was ineffectual. Even by the intracerebral route, however, it was not possible to counteract LSD-25 in all the animals. Detailed data showing the antagonism between serotonin and LSD-25 in isolated segments of carotid arteries were recorded. The behavioral change caused in mice by LSD-25 was viewed as resulting from a cerebral deficiency of serotonin induced by this antimetabolite.

* With the technical assistance of G. Schaffner and E. Van Winkle.

[1] G. Barger, *Ergot and Ergotism* (London: Gurney & Jackson, 1931); W. A. Stoll, *Schweiz. Arch. Neurol. Psychiat.*, **68**, 279, 1947.

[2] D. W. Woolley and E. Shaw, *Federation Proc.*, **12**, 542, 1953.

[3] E. Shaw and D. W. Woolley, *J. Biol. Chem.*, **203**, 979, 1953.

[4] J. H. Gaddum, C. O. Hebb, A. Silver, and A. A. B. Swan, *Quart. J. Exptl. Physiol.*, **38**, 255, 1953.

[5] D. W. Woolley and E. Shaw, these PROCEEDINGS, **40**, 228, 1954.

[6] D. W. Woolley and E. Shaw, *Brit. Med. J.*, **2**, 122, 1954.

[7] LSD-25 was kindly supplied by Sandoz Chemical Works, Inc.

[8] D. W. Woolley and E. Shaw, *J. Biol. Chem.*, **203**, 69, 1953.

[9] M. C. Greig and M. K. Carter, *Arch. Biochem. and Biophys.*, **52**, 175, 1954.

[10] M. Roche e Silva and J. Ribiero do Valle, *Abstr. XIX Intern. Physiol. Congr.* (1953), p. 738.

[11] H. A. Abramson and L. T. Evans, *Science*, **120**, 990, 1954.

FINE STRUCTURE OF A GENETIC REGION IN BACTERIOPHAGE

BY SEYMOUR BENZER

BIOPHYSICAL LABORATORY, DEPARTMENTS OF BIOLOGICAL SCIENCES AND PHYSICS, PURDUE UNIVERSITY, LAFAYETTE, INDIANA

Communicated by M. Delbrück, April 6, 1955

This paper describes a functionally related region in the genetic material of a bacteriophage that is finely subdivisible by mutation and by genetic recombination. The group of mutants resembles similar cases which have been observed in many organisms, usually designated as "pseudo-alleles." (See reviews by Lewis[1] and

Pontecorvo.[2]) Such cases are of special interest for their bearing on the structure and function of genetic determinants.

The phenomenon of genetic recombination provides a powerful tool for separating mutations and discerning their positions along a chromosome. When it comes to very closely neighboring mutations, a difficulty arises, since the closer two mutations lie to one another, the smaller is the probability that recombination between them will occur. Therefore, failure to observe recombinant types among a finite number of progeny ordinarily does not justify the conclusion that the two mutations are inseparable but can only place an upper limit on the linkage distance between them. A high degree of resolution requires the examination of very many progeny. This can best be achieved if there is available a selective feature for the detection of small proportions of recombinants.

Such a feature is offered by the case of the rII mutants of T4 bacteriophage described in this paper. The wild-type phage produces plaques on either of two bacterial hosts, B or K, while a mutant of the rII group produces plaques only on B. Therefore, if a cross is made between two different rII mutants, any wild-type recombinants which arise, even in proportions as low as 10^{-8}, can be detected by plating on K.

This great sensitivity prompts the question of how closely the attainable resolution approaches the molecular limits of the genetic material. From the experiments of Hershey and Chase,[3] it appears practically certain that the genetic information of phage is carried in its DNA. The amount of DNA in a particle of phage T2 has been determined by Hershey, Dixon, and Chase[4] to be 4×10^5 nucleotides. The amount for T4 is similar.[5] If we accept the model of DNA structure proposed by Watson and Crick,[6] consisting of two paired nucleotide chains, this corresponds to a total length of DNA per T4 particle of 2×10^5 nucleotide pairs. We wish to translate linkage distances, as derived from genetic recombination experiments, into molecular units. This cannot be done very precisely at present. It is not known whether all the DNA in a phage particle is indispensable genetic material. Nor is it known whether a phage "chromosome" (i.e., the physical counterpart of a linkage group identified by genetic means) is composed of a single (duplex) DNA fiber or whether genetic recombination is equally probable in all chromosomal regions. For the purpose of a rough calculation, however, these notions will be assumed to be true. Thus we place the total linkage map of T4 in correspondence with 2×10^5 nucleotide pairs of DNA. The total known length of the three linkage groups[7] in phage T4 amounts to some 100 units (one unit = 1 per cent recombination in a standard cross). In addition, there is evidence[8] for roughly another 100 units of length connecting two of the groups. Therefore, if we assume 200 recombination units to correspond to 2×10^5 nucleotide pairs, the recombination per nucleotide pair is 10^{-3} per cent. That is to say, given two phage mutants whose mutations are localized in their chromosomes at sites only one nucleotide pair apart, a cross between these mutants should give rise to a progeny population in which one particle in 10^5 results from recombination *between* the mutations (provided, of course, that recombination is possible between adjacent nucleotide pairs). This computation is an exceedingly rough one and is only intended to indicate the order of magnitude of the scale factor. Some preliminary results are here presented of a program designed to extend genetic studies to the molecular (nucleotide) level.

r Mutants.—The wild-type phages T2, T4, and T6 produce small plaques with rough edges when plated on strain B of *Escherichia coli*. From sectors of clearing in these plaques, mutants can be readily isolated which produce large, sharp-edged plaques (Hershey[9]). These mutants have been designated "r" for rapid lysis; they differ from the wild type by a failure to cause "lysis inhibition" on strain B (Doermann[10]). The wild type has a selective advantage over *r* mutants when the two types grow together on B. The genetics of r mutants was studied by Hershey and Rotman,[11] who found three regions in the linkage map of T2 in which various mutations causing the r phenotype were located, including one large "cluster" of mutants which were shown to be genetically distinct from one another. The genetic study of T4 by Doermann and Hill[7] showed r regions corresponding to two of those in T2. T6 also has at least two such r regions.

The rII Group.—For all three phages, T2, T4, and T6, the r mutants can be separated into groups on the basis of their behavior on strains other than B. This paper will be concerned only with one group, which will be called the "rII group." Mutants of the rII group are distinguished from those of other groups, and from wild type, by a failure to produce plaques on certain lysogenic strains[12] of *E. coli* which carry phage λ. As shown in Table 1, a mutant of the rII group produces

TABLE 1

PHENOTYPES (PLAQUE MORPHOLOGY) OF T4 WILD AND rII MUTANT PLATED ON VARIOUS HOSTS

	Host Strain		
	E. coli B	*E. coli* K12S	*E. coli* K12S (λ)
T4 wild type	Wild	Wild	Wild
T4 rII mutant	r Type	Wild	. .

r-type plaques on strain B, wild-type plaques on strain K12S (nonlysogenic strain sensitive to λ), and no plaques on K12S (λ) (derived from K12S by lysogenization with λ). The wild-type phage produces similar plaques on all three strains. In the case of T4, with which we shall be concerned in this paper, the efficiencies of plating are approximately equal on the three strains, except, of course, for rII on K12S (λ). The three bacterial strains will be here designated as "B," "S," and "K."

Approximately two-thirds of the independently arising r mutants isolated on B are of the rII type. This group includes the "cluster" of r mutants of T2 described by Hershey and Rotman and the r47 and r51 mutants described by Doermann and Hill in the corresponding map region of T4 but does not include r mutants located outside that region. Similarly, all newly isolated mutants showing the rII character have turned out to fall within the same region, as indicated in Figure 1.

The properties of the rII group are especially favorable for detailed genetic study. An rII mutant has three different phenotypes on the three host strains (Table 1): (1) altered plaque morphology on B, (2) indistinguishable from wild type on S, and (3) unable to produce plaques on K. These properties are all useful. By virtue of their altered plaque type on B, r mutants are readily isolated, and those of the rII group are identified by testing on K. Where it is desired to avoid a selective disadvantage compared with wild type, e.g., in measuring mutation

rates, S can be used as a nondiscriminating host. The failure of rII mutants to plate on K enables one to detect very small proportions of wild-type particles due to reversion or due to recombination between different rII mutants.

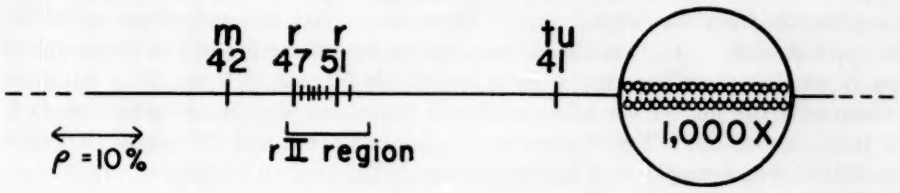


FIG. 1.—Partial linkage map of T4 (Doermann), indicating the location of the rII region. *m* and *tu* designate "minute plaque" and "turbid plaque" mutations. The circular inset shows, diagrammatically, the corresponding dimensions of the DNA chain magnified 1,000 diameters.

Fate of rII Mutants in K.—Wild-type and rII mutants adsorb equally well to strains S and K. Whereas the wild type provokes lysis and liberation of a burst of progeny on both strains, the rII mutant grows normally only on S. Infection of K with an rII mutant provokes very little (and/or very late) lysis, although all infected cells are killed. The block in growth of rII mutant is associated with the presence of the carried phage λ. The reason for this association is unknown.

Quantitative Differences in Phenotype.—While all rII mutants show the same phenotypic effect of poor multiplication on K, they differ in the degree of this effect. A certain proportion of K infected with rII actually liberates some progeny, which can be detected by plating the infected cells on B. The fraction of infected cells yielding progeny defines a "transmission coefficient" characteristic of the mutant. The transmission coefficient is insensitive to the multiplicity of infection but depends strongly upon the physiological state of the bacteria (K) and upon temperature. Under given conditions, however, the coefficient can be used as a comparative index of degree of phenotypic effect, a "leaky" mutant having a high coefficient. As can be seen in Table 2, a wide range of values is found.

TABLE 2

PROPERTIES OF T4 MUTANTS OF THE rII GROUP*

Mutant Number	Map Position	Transmission Coefficient	Reversion Index (units of 10^{-6})
r47	0	0.03	<0.01
r104	1.3	.91	<1
r101	2.3	.03	4.5
r103	2.9	.02	<0.2
r105	3.4	.02	1.8
r106	4.9	.55	<1
r51	6.7	.02	170
r102	8.3	.02	<0.01

* Three parameters are given for each mutant. The map position is computed from the sum of the nearest intervals shown in Figure 2 and is given in percentage recombination units, taking the position of r47 as zero. The "transmission coefficient" is a measure of phenotypic effect determined by infecting bacteria K with the mutant in question and is given as the fraction of such infected cells yielding plaques on strain B. The "reversion index" is the average fraction of wild-type particles arising in lysates of the mutant grown from a small inoculum on a nonselective host.

Plaques on K.—Some rII mutants produce no plaques on K, even when as many as 10^8 particles (as measured by plaque count on B) of a stock are plated. Other rII mutants, however, produce various proportions of plaques on K. When the

plaques appearing on K are picked and retested, they fall into three categories: (1) a type which, like the original mutant, produces very few plaques on K and r-type plaques on B; (2) a type which produces plaques (often smaller than wild type) on K with good efficiency but r-type plaques on B; and (3) a type indistinguishable from the original wild. These three types are understood to be due to the following: (1) "leaking" effects, i.e., ability of the mutant to grow slightly on K, so that there is a chance for a few visible plaques to form; (2) a mutation which partially undoes the effect of the rII mutation, so that multiplication in K is possible, but the full wild phenotype is not achieved; and (3) apparent reverse mutation, which may or may not be genuine, to the original wild type.

The proportion of each type occurring in a stock is characteristic and reproducible for a particular rII mutant but differs enormously from one rII mutant to another. There is no evident correlation in the rates of occurrence of the three types.

Reversion Rates of rII Mutants.—Reversion of r mutants to a form indistinguishable from wild type was demonstrated by Hershey,[9] who made use of the selective advantage of wild type on B to enrich its proportion in serial transfers. Given the inability of rII mutants to produce plaques on K, such reversions are easily detected, even in very small proportion. An index to the frequency of reversion of a particular rII mutant can be obtained by preparing a lysate from a small inoculum (about 100 particles, say, so that there is very little chance of introducing a wild-type particle present in the stock). If S is used as the host, both rII mutant and any reversions which arise can multiply with little selection, as shown by control mixtures. The average fraction of wild-type particles present in several lysates is an index which can be shown to be roughly proportional to the probability of reversion per duplication of the rII mutant. Under the conditions of measurement the index is of the order of 10–20 times the probability of reversion per duplication. The plaques appearing on K must be tested by picking and replating on B. This eliminates the "spurious" plaques produced by partial reversions and by leaky mutants, which show up as r type on B. As may be seen in Table 2, the reversion indices for rII mutants vary over a very wide range. One mutant has been found which reverts 10 times more frequently than r51, so that the reversion rates cover a known range of over 10^5-fold.

It has not been proved that these apparent reversions constitute a genuine return to the original wild type. However, the possibility of suppressor mutations distant from the site of the rII mutation has been ruled out by backcrosses to the original wild type. Krieg[13] found very few, if any, r-type recombinants in backcrosses of several reversions, localizing the reverse changes to within a few tenths of a per cent linkage distance from the original rII mutations. One case of "partial reversion" has also been tested by backcrossing, and failure to observe rII-type recombinants localized the "partial reverse mutation" to within the rII region.

Mapping of the rII Region.—A cross between two rII mutants is made by infecting a culture of B with equal multiplicities (three per bacterium) of each type. The yield after lysis contains the two parental types and, if the parents are genetically distinct, two recombinant types, the double mutant and wild type. In the average yield from many cells, the recombinant types occur in equal numbers.[11] In all cases thus far tested, double rII mutants, like single mutants, do not produce plaques on K. On the assumption that this is generally true, the proportion of

recombinants in the yield can be measured simply by doubling the ratio of the plaque count on K (which registers only the wild recombinant) to the count on B (which registers all types). The percentage of wild type thus measured agrees well with a direct count of plaque types on B.

In this way, a series of six rII mutants of T4 (the first six isolated—not selected in any way) have been crossed with each other and with r47 and r51 (kindly supplied by A. H. Doermann) in 23 of the 28 possible pairs. The results of these crosses are given in Figure 2 and are compatible with the indicated seriation of the mutants. The distances are only roughly additive; there is some systematic deviation in the sense that a long distance tends to be smaller than the sum of its component shorter ones. Part of this discrepancy is accounted for by the Visconti-Delbrück correction for multiple rounds of mating.[14] Reversion rates were small enough to be negligible in these crosses. Thus, while all rII mutants in this set fall into a small portion of the phage linkage map, it is possible to seriate them unambiguously, and their positions *within* the region are well scattered.

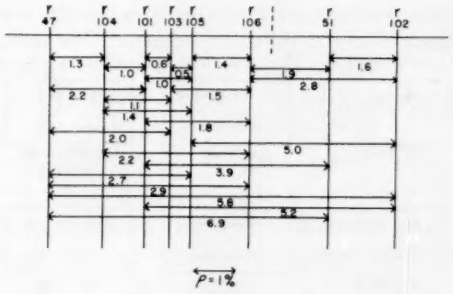


FIG. 2.—Larger-scale map of eight rII mutants, including Doermann's r47 and r51. Newly isolated mutants are numbered starting with *101*. The recombination value (in per cent) for each cross is obtained by plating the progeny on K and on B and doubling the ratio of plaque count on K to count on B.

Tests for Pseudo-allelism.—The functional relatedness of two closely linked mutations causing similar defects may be tested by constructing diploid heterozygotes containing the two mutations in different configurations.[1, 2] The *cis* form, with both mutations in one chromosome, usually behaves as wild type, since the second chromosome supplies an intact functional unit (or units). However, the *trans* form, containing one of the mutations in each chromosome, may or may not produce the wild phenotype. If it does, it is concluded that the two mutations in question are located in separate functional units.

In applying this test to the rII mutants, the diploid heterozygote can be simulated by a mixed infection with two kinds of phage. The rII phenotype is a failure to lyse K, whereas the wild phenotype is to cause lysis. If K is mixedly infected with wild type and rII mutant, the cells lyse, liberating both types of phage. Thus the presence of wild type in the cell supplies the function which is defective in rII type, and the rII mutation can be considered "recessive." Although it has not yet been tested, the *cis* configuration of double rII mutant plus wild type is also presumed to produce lysis in all cases. The *trans* configuration is obtained by infecting K with the pair of rII mutants in question. This is found to give lysis or not, depending upon which rII mutants compose the pair. The results are summarized by the dotted line in Figure 2, indicating a division of the rII region into two segments. If both mutants belong to the same segment, mixed infection of K gives the mutant phenotype (very few cells lyse). If the two mutants belong to different segments, extensive lysis occurs with liberation of both infecting types (and recombinants). These results are summarized in Figure 3. Thus, on the basis of this

test, the two segments of the rII region correspond to independent functional units.

Actually, for mixed infection of K with two (nonleaky) mutants of the *same* segment, a very small proportion of the cells do lyse and liberate wild recombinants, that proportion increasing with the linkage distance between the mutations. For two rII mutants separated by 1 per cent linkage distance (measured by a standard cross on B) the proportion of mixedly infected K yielding any wild particles is about 0.2 per cent.

This value has bearing upon the effect upon K/B values of the heterozygous phage particles which arise in a cross between two rII mutants on B. In such a cross between closely linked rII mutants, the progeny should include about 2 per cent of particles containing a *trans* configuration heterozygous piece.[15] When one of these is plated on K, there is a certain chance that a wild recombinant may form in the first cycle of infection, leading to production of a plaque. If it is assumed that these are no more likely to do so than a mixed infection of K with two complete mutant particles, it can be concluded that the effect of these heterozygous particles upon the count on K is negligible, provided that both rII mutants belong to the same segment. For mutants in different segments, however, the "efficiency" of the heterozygous particles should be much greater, and recombination values measured by the K/B method should run considerably higher than the true values. The recombination values in Figure 2 for crosses which transgress the segmental divide are probably subject to some correction for this reason.

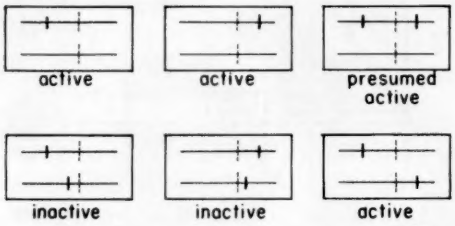


FIG. 3.—Summary of tests for "position-effect pseudo-allelism" of rII mutants. Each diagram represents a diploid heterozygote as simulated by mixed infection of a bacterium (K) with two types of phage containing the indicated mutations. *Active* means extensive lysis of the mixedly infected cells; *inactive* means very little lysis. The dotted line represents a dividing point in the rII region, the position of which is defined by these results.

Rough Mapping by Spot Test.—If a stock of either of two rII mutants is plated on K, no plaques arise; but if both are plated together, some bacteria become infected by both mutants and, if this leads to the occurrence of wild-type recombinants, plaques are produced. If the two mutants are such that wild recombinants cannot arise between them (e.g., if they contain identical mutations), no plaques appear. A given rII mutant may thus be tested against several others on a single plate by first seeding the plate with K plus the mutant in question (in the usual soft agar top layer) and then spotting with drops containing the other rII mutants.

Inspection of such a plate immediately places the unknown mutant in the proper segment, since spotting any mutant of segment A against any mutant of segment B gives a very clear spot, due to the extensive lysis of mixedly infected bacteria. However, for a pair of mutations belonging to the same segment, plaques are produced only by the relatively few mixedly infected bacteria which give rise to wild recombinants. The greater the linkage distance between the mutations, the larger the number of plaques that appear in the spot. A group of mutants of the same segment may thus be seriated by seeding one plate with each and spotting with all the others. Given a previously seriated group, a new mutant can thus be quickly

located within the group. This method works best for mutants which are stable (i.e., low reversion rate) and nonleaky, so that large numbers of phage particles can be plated. Reversions or pronounced leaking effects obviously cause an obscuring background.

This test has been applied to a large group of stable, nonleaky rII mutants. Their approximate locations as deduced from these tests are shown in Figure 4. Some of the mutants showed anomalies which made it impossible to locate them as members of a series. They gave very little recombination with any of the mutants located within a certain span, while behaving normally with respect to mutants located outside that span. They are indicated in Figure 4 by horizontal lines extending over the span.

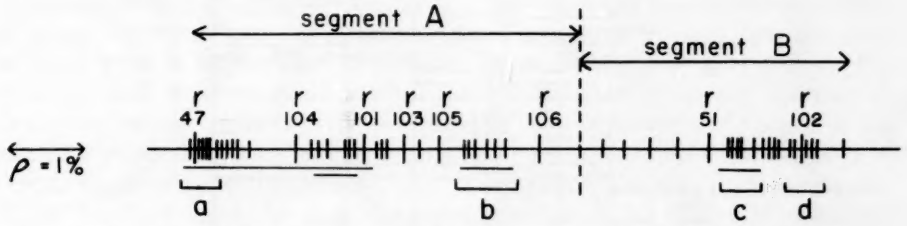


FIG. 4.—Preliminary locations of various rII mutants, based upon spot tests

Spot tests on numerous other mutants have shown that mutants of varied reversion rates, transmission coefficients, and rates of "partial reversion" occur at scattered positions in both segments.

Mapping of "Microclusters."—The spot test enables us to pick out "microclusters," i.e., groups of very closely neighboring mutations. Four such groups selected for further study are indicated in Figure 4, and the results of mapping them are given in Figure 5. While some intervals show reasonably good additivity properties, there are some mutants which give violently anomalous results. Thus in microcluster *a*, r47 gives no wild recombinants (i.e., less than 1 in 10^6) with any of the other three mutants, but two pairs of the three do show recombination. These results can be understood if it is assumed that each mutation extends over a certain length of the chromosome, and production of wild type requires recombination within the space *between* those lengths. According to this interpretation, the mutations would cover the lengths indicated by the bars in Figure 5. These anomalies resemble those observed in the spot tests, only they are more limited in span.

This observation raises the question of whether there exist true "point" mutations (i.e., involving an alteration of only one nucleotide pair) or whether all mutations involve more or less long pieces of the chromosome. It must be remembered that the mutants used in these experiments were selected for extreme stability against reversion. This procedure would be expected to enrich the proportion of mutants containing gross chromosomal alterations. So far as is known, the anomalous cases observed could equally well be imagined to be due to double (i.e., two near-by "point") mutations, inversions, or deletions of the wild-type chromosome. In continuing these experiments, it would seem well advised to employ only mutants for which some reversion is observed.

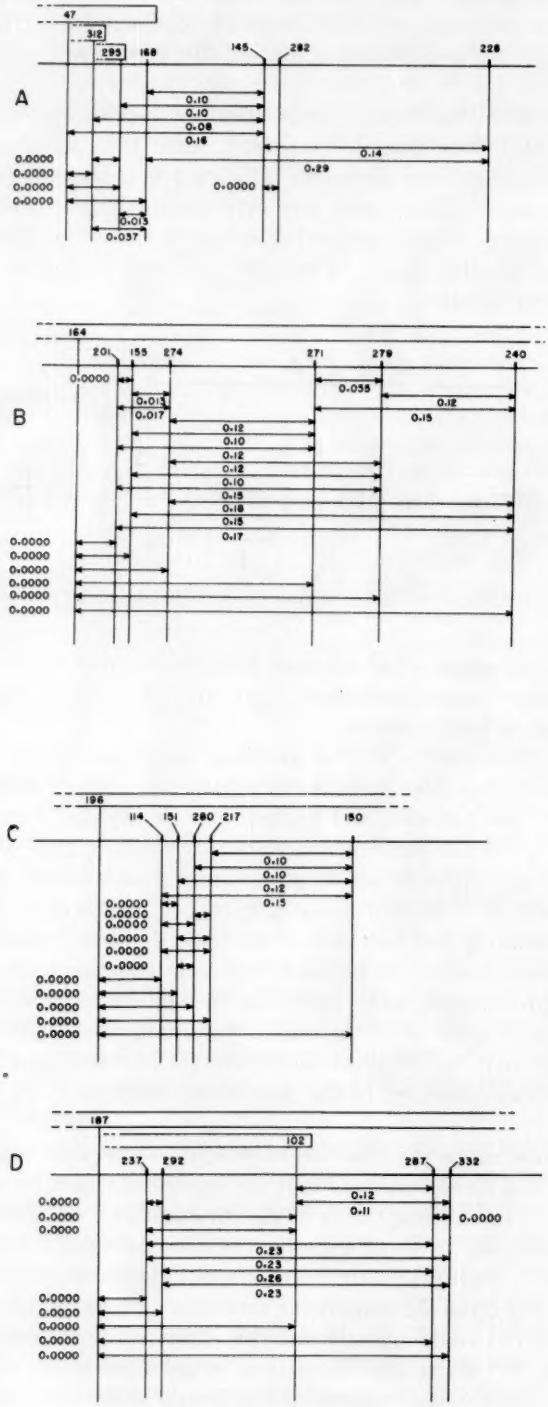


FIG. 5.—Maps of microclusters

Discussion.—The set of rII mutants defines a bounded region of a linkage group in which mutations may occur at various locations, all the mutations leading to qualitatively similar phenotypic effects. The rII region would seem, therefore, to be functionally connected, so that mutations arising anywhere within the region affect the same phenotype. This effect is expressed, in case strain B is the host, by failure to produce lysis inhibition; in case S is the host, by no consequence; and in case K is the host, by inability to multiply normally. The failure of an rII mutant to mature in K can be overcome by the presence of a wild-type phage in the same cell. This could be understood if the function of the region in the wild-type "chromosome" were to control the production of a substance or substances needed for reproduction of this phage in K cells.

The phenotypic test for "pseudo-allelism" leads to the division of the region into two functionally distinguishable segments. These could be imagined to affect two necessary sequential events or could go to make up a single substance the two parts of which must be unblemished in order for the substance to be fully active. For example, each segment might control the production of a specific polypeptide chain, the two chains later being combined to form an enzyme. While it is not known whether this sort of picture is applicable, a model of this kind is capable of describing the observed properties of the rII mutants. The map position of a mutation would localize a change in the region (and also in the "enzyme" molecule), the reversion rate would characterize the type of change involved in the genetic material, and the degree of phenotypic effect would be an expression of the degree of resultant change in the activity of the enzyme. A "leaky" mutant would be one where this latter effect was small. While no obvious correlation has yet been observed among these three parameters of rII mutants, one may well show up upon more exhaustive study.

"Clustering" of similar mutants separable by crossing-over has been observed for several characters in phage by Doermann and Hill and appears to represent the rule. This may well be the rule in all organisms, simply because functional genetic units are composed of smaller recombinational and mutational elements. One would expect to see this effect more readily in phage because the probability of recombination per unit of hereditary material is much greater than for higher organisms.

By extension of these experiments to still more closely linked mutations, one may hope to characterize, in molecular terms, the sizes of the ultimate units of genetic recombination, mutation, and "function." Our preliminary results suggest that the chromosomal elements separable by recombination are not larger than the order of a dozen nucleotide pairs (as calculated from the smallest non-zero recombination value) and that mutations involve variable lengths which may extend over hundreds of nucleotide pairs.

In order to characterize a unit of "function," it is necessary to define what function is meant. The entire rII region is unitary in the sense that mutations anywhere within the region cause the rII phenotype. On the basis of phenotype tests of *trans* configuration heterozygotes, this region can be subdivided into two functionally separable segments, each of which is estimated to contain of the order of 4×10^3 nucleotide pairs. If one assumes that each segment has the "function" of specifying the sequence of amino acids in a polypeptide chain, then the specification of each individual amino acid can as well be considered a unitary function. It

would seem feasible, with this system, to extend genetic studies even to the level of the latter functional elements.

Summary.—It has been discovered that the mutations in the rII region of phage T4 have a characteristic in common which sets them apart from the mutations in all other parts of the map. This characteristic is a host-range reduction, namely, a failure to produce plaques on a host (K) lysogenic for phage λ. The mutant phage particles adsorb to and kill K, but normal lysis and phage release do not occur.

All mutants with this property are located within a sharply defined portion of the phage linkage map. Within that region, however, their locations are widely scattered. An unambiguous seriation of the mutants, with roughly additive distances, can be accomplished, except for certain anomalous cases.

The simultaneous presence of a wild-type phage particle in K enables the multiplication of rII mutants to proceed, apparently by supplying a function in which the mutant is deficient. A heterozygous diploid in the *trans* configuration is simulated by a mixed infection of K with two mutant types. The application of the phenotype test to pairs of rII mutants leads to the division of the region into two functionally separable segments.

Spontaneous reversion to wild-type had been observed for most of these mutants. It remains to be seen whether these are genuine reversions. Each mutant reverts at a characteristic rate, but the rates for different mutants differ enormously. Partial reversions to intermediate types are also observed.

The mutants differ greatly in degree of residual ability to grow on K. There is no evident correlation between map position, reversion rate, and degree of residual activity of the various mutants.

The selective feature of K for wild-type recombinants offers the possibility of extending the recombination studies to an analysis of the fine details of the region.

Preliminary studies of this type indicate that the units of recombination are not larger than the order of one dozen nucleotide pairs and that mutations may involve various lengths of "chromosome."

I am much indebted to A. D. Hershey and A. H. Doermann for stocks of their genetically mapped mutants, to Sydney Brenner and David Krieg for stimulating discussion, and to Max Delbrück for his invaluable moderating influence.

* Supported by a grant-in-aid from the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council.

[1] E. B. Lewis, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 159–174, 1951.

[2] G. Pontecorvo, *Advances in Enzymol.*, **13**, 121–149, 1952.

[3] A. D. Hershey and M. Chase, *J. Gen. Physiol.*, **36**, 39–56, 1952.

[4] A. D. Hershey, J. Dixon, and M. Chase, *J. Gen. Physiol.*, **36**, 777–789, 1953.

[5] E. K. Volkin, personal communication.

[6] J. D. Watson and F. H. C. Crick, *Cold Spring Harbor Symposia Quant. Biol.*, **18**, 123–131, 1953.

[7] A. H. Doermann and M. B. Hill, *Genetics*, **38**, 79–90, 1953.

[8] G. Streisinger and V. Bruce, personal communication.

[9] A. D. Hershey, *Genetics*, **31**, 620–640, 1946.

[10] A. H. Doermann, *J. Bacteriol.*, **55**, 257–276, 1948.

[11] A. D. Hershey and R. Rotman, *Genetics*, **34**, 44–71, 1949.

[12] E. M. Lederberg and J. Lederberg, *Genetics*, **38**, 51–64, 1953.

[13] D. Krieg, personal communication.

[14] N. Visconti and M. Delbrück, *Genetics*, **38**, 5–33, 1953.

[15] A. D. Hershey and M. Chase, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 471–479, 1951; C. Levinthal, *Genetics*, **39**, 169–184, 1954.

*CHROMOSOMAL ABERRATIONS IN DATURA DUE TO VARIOUS KINDS OF IRRADIATION**

By Jean M. Cummings, Lewis Goldstein, and A. F. Blakeslee†

WESTERN RESERVE UNIVERSITY, CLEVELAND, OHIO; SMITH COLLEGE, NORTHAMPTON, MASSACHUSETTS; AND GENETICS EXPERIMENT STATION, SMITH COLLEGE, NORTHAMPTON, MASSACHUSETTS

Communicated by M. Demerec, April 19, 1955

Recent work on the biological effects of neutron radiation has been aimed to a great extent at determining the relative biological efficiency of these radiations and at finding out whether the results with neutrons differ qualitatively from the results with other types of radiations. As part of this work, a chromosomal study was initiated, in 1952, in *Datura stramonium*, Line 1, whose seeds had been treated with thermal neutrons.[1] This study was expanded in 1953 to include gamma and X-rays as well as neutrons from a nuclear detonation.[2] The present work is an extension, dealing with chromosomal aberrations due to neutrons from a cyclotron, as well as from a nuclear detonation, from a neutron flux in the Brookhaven pile, and to X-rays.

The earlier work had indicated that thermal neutrons might be qualitatively different in their effect on *Datura*,[1] and it was hoped that further testing would clarify this. It was also felt that a comparison of detonation neutrons and cyclotron neutrons should be made in this plant, even though work on other test organisms had indicated that the source of neutrons apparently had little effect on the results.[3]

Experimental Procedures and Results.—The treatment of the seeds with the various radiations has been described in the paper on pollen abnormalities in these plants.[4]

In sampling, the buds were taken at random from the plants and examined, and slides were made of any anthers showing division figures. The slides were scored at metaphase and at anaphase for aberrations, and configurations were usually checked by at least one, and sometimes by two, workers in addition to the finder. As part of the policy of the laboratory, no general comparison of results was made until after the work was almost completed. The iron-acetocarmine technique was used, and segmental interchanges, large deletions, and large duplications could be observed. Earlier stages in meiosis cannot be studied in this plant by this method, so that small duplications and deficiencies, as well as inversions, could not be seen.

The thermal-neutron dosage unfortunately proved to be too severe. Although 149 of the 300 treated seeds germinated, only 16 of them grew large enough to permit buds to be examined, and only 14 gave usable slides. So far as could be determined with such a small sample, the types of aberrations found were the same as those caused by other kinds of radiations (Table 2). There may have been some difference, however, in the frequency of the various types of aberrations in these plants.

† Died November 16, 1954.

As can be seen from Table 1, the percentages of chromosomal aberrations found with X-rays and with the detonation neutrons compare well for the two years. The percentages of aberration were somewhat lower in 1954, but, considering the great differences in the numbers of plants examined in the two years, these differences were slight. In the case of the X-rays, in 1953, 23 plants were examined, and 4 aberrations were found.[2] In 1954 the figures were, respectively, 104 and 16 (Table 2). For detonation neutrons at 760 rep, the figures were 56 plants tested, with 12 showing aberrations, in 1953 and 104 plants tested, with 21 having chromosomal abnormalities, in 1954. In the two cases the finding or not finding of a single abnormal plant, in 1953, would have changed the percentage points by about 2 per cent for the detonation result and by about 4 per cent for the X-ray result.

TABLE 1

COMPARATIVE EFFECTS OF RADIATIONS IN DIFFERENT YEARS BASED ON PERCENTAGE OF ABERRATIONS

EXPOSURE	DOSE	RESULTS (PER CENT) 1953	1954
X-rays	10,000 r	17	15
Detonation	2,000 rep	56	..
Detonation	1,400 rep	34	..
Detonation	760 rep	21	20
Detonation	246 rep	12	..
Cyclotron	650 rep	..	2.6
Thermal neutrons	7.09×10^8 cm^2/sec (3 hr)	..	64

TABLE 2

COMPARISON OF ABERRATION TYPES FROM DIFFERENT RADIATIONS

Exposure	No. Plants	No. Aberr.	⊙4	Chain 4	Necktie	Def.	Dup.	⊙6+ ⊙4	Necktie + ⊙4	Unequal Bivalent
X-rays, 10,000 r	104	16	9	2	1	2	..	..	1	1
Detonation, 760 rep	104	21	12	3	..	3	2	..	..	1
Cyclotron, 650 rep	75	2	..	..	1	1	..	..	..	..
Thermal neutrons, 7.09 × 10^8 cm^2/sec (3 hr)	14	9	4	1	..	..	1	3	..	..
Total	297	48	25	6	2	6	3	3	1	2

Table 1 also shows the results obtained with cyclotron neutrons. It is quite apparent here that there has been a great quantitative difference in response of the chromosomes to detonation and cyclotron neutrons. The actual number of plants, grown from cyclotron-treated seeds, that were examined was 75, and only 2 showed any chromosomal abnormalities. It should perhaps be mentioned that, of the 75 thermal-neutron plants analyzed, one author examined 40 and found 2 aberrations, and the second author examined 35 and found no abnormalities. In view of this, and in view of the relatively large number of plants examined, the possibility that any significant number of gross chromosomal abnormalities was overlooked is extremely slight.

Also of some interest in connection with these results is the fact that there were 12 per cent of morphologically aberrant plants (40 out of 333) among those treated with detonation neutrons and only 2 per cent (5 out of 254) among those treated with cyclotron neutrons.

Table 2 shows the various types and the frequencies of aberrations found in 1954 with the different kinds of radiations. The order of frequency is as follows: ring of four, chain of four, deficiency, duplication, ring of four plus ring of six, necktie, unequal bivalent, and necktie plus ring of four. As is usually true in irradiated plants at metaphase, rings of four are by far the most common type of interchange found, constituting in this series more than 50 per cent of the aberrations.

The deficiencies mentioned in the seventh column of Table 2 are of various kinds and degrees and were classified together under this heading for convenience. For example the two X-ray deficiencies were deficiencies of a whole or a large portion of a translocated chromosome resulting in metaphase plates showing 10 bivalents plus a chain of three rather than the normal complement of 12 bivalents. The three detonation deficiencies included one plant with 11 bivalents plus a univalent; one plant with 10 bivalents, a chain of three, and a small free fragment; and, finally, one plant with 11 bivalents, a univalent, and a small free fragment. The cyclotron deficiency was the loss of a whole chromosome, resulting in 11 bivalents and a univalent at metaphase. The duplications were extra fragments of chromosomes present with 12 bivalents. The unequal bivalents might be either duplications or deficiencies and therefore have been included under a separate heading.

Discussion.—Yost in 1953 found some indication that rings of four were constantly associated with 50 per cent aborted pollen in thermal-neutron-treated plants of *Datura*. This could not be substantiated in the present work, because, unlike barley, in which germination of thermal-neutron-treated seeds and maturation of the plants are affected only slightly,[5] *Datura* seeds are quite susceptible to this type of radiation, as shown by the comparatively low germination rate and extremely high mortality of seedlings. It was found, however, that four of the five plants which had 12 normal bivalents also had 50 per cent or more aborted pollen, which may indicate that the association of bad pollen and chromosome rings was merely fortuitous. This seems particularly plausible in view of the fact that, in *Datura*, rings of four are not associated with pollen abortion when these interchanges occur in nature or when they are induced by aging of the seeds,[6] by gamma rays, X-rays or fast neutrons.[2]

Although the types of chromosomal abnormalities found this year in the thermal-neutron-treated plants were the same as those in plants treated with other kinds of radiation, the frequency of ring of four plus ring of six (Table 2) was much greater in these plants. This was thought to be due to the greater effect of the given dosage of thermal neutrons. However, Caldecott[7] also found a larger percentage of rings of four plus rings of six in thermal-neutron-treated barley, and this is a plant apparently not damaged to any extent by high dosages of thermal neutrons. In neither case have sufficient plants been analyzed to determine whether or not this difference is statistically significant, but the similarity in results with the two plants is interesting.

The results with the X-rays and detonation neutrons reaffirm the fact that these neutrons are much more effective in causing chromosomal abnormalities than are the X-rays. This, of course, is in line with the results obtained by many investigators on various organisms.[3, 8, 9]

In view of the great divergence of results with detonation and cyclotron neutrons in *Datura*, it is extremely unfortunate that a series of dosages with the cyclotron neutrons was not run for comparison with the results from the detonation. The percentages of chromosomal aberrations from different dosages of detonation neutrons which were determined in 1953 are shown in Table 1, and, as can be seen, they increase about linearly with dose. The results with the cyclotron neutrons cannot be fitted in with the results from the detonation neutrons. It is true that there was an unknown amount of gamma radiation accompanying the detonation neutrons, but there was at least 10 per cent gamma radiation contaminating the dosage of cyclotron neutrons also. The result is even more peculiar when account is taken of the fact that in other organisms studied there has been good agreement in results of similar dosages between cyclotron and detonation neutrons.[3, 8, 10] Some discrepancies have been noted, but nothing of the magnitude of difference presented here.[11,12,13] One cannot, of course, expect perfect correlation with results found in other experiments: the organisms differ; the treatment, in many cases, has been done at different places with different instruments; and the dosages have been measured by different methods and with various kinds of measuring devices. Even taking these factors into account, there is still such a large discrepancy in response here that it can hardly be explained by the above. Resolution of this problem must await a more comprehensive testing of neutron treatments of seeds. As yet, little work has been done on chromosomal abnormalities in plants grown from neutron-irradiated seed. Accumulation of this type of information will undoubtedly produce fundamental knowledge about the biological damage caused by neutrons.

Summary.—X-rays, thermal neutrons, and fast neutrons from a nuclear detonation and from a cyclotron cause qualitatively similar chromosomal aberrations in *Datura*. Thermal neutrons are very damaging to *Datura* and may cause a higher frequency of rings of four plus rings of six than is caused by other radiations. Cyclotron neutrons were not as effective in causing chromosomal aberrations as were neutrons from a nuclear detonation. The reason for this is unknown.

The authors are indebted to the staffs of the Oak Ridge and Brookhaven National Laboratories, who treated the seeds. Special thanks are due to Dr. C. W. Sheppard for his aid.

* Contributions from the Department of Botany, Smith College, new ser., No. 64. This work was supported by contract with the United States Atomic Energy Commission.

[1] H. T. Yost, W. R. Singleton, and A. F. Blakeslee, these PROCEEDINGS, **39**, 292–297, 1953.

[2] H. T. Yost, J. M. Cummings, and A. F. Blakeslee, these PROCEEDINGS, **40**, 447–451, 1954.

[3] H. H. Plough, *et al.*, Operation Upshot/Knothole, Projects 23.4–23.14 and 23.16, United States Atomic Energy Commission, January, 1954.

[4] J. L. Spencer and A. F. Blakeslee, these PROCEEDINGS, **41**, 307–312, 1955.

[5] R. S. Caldecott, E. F. Frolik, and R. Morris, these PROCEEDINGS, **38**, 804–809, 1952.

[6] J. L. Cartledge and A. F. Blakeslee, these PROCEEDINGS, **20**, 103–110, 1934.

[7] R. S. Caldecott, B. H. Beard, and C. O. Gardner, *Genetics*, **39**, 240–259, 1954.

[8] A. D. Conger, *Am. Naturalist*, **88**, 215–224, 1954.

[9] N. H. Giles, F. J. de Serres, and A. V. Beatty, *Genetics*, **38**, 416–421, 1953.

[10] J. S. Kirby-Smith and C. P. Swanson, *Science*, **119**, 42–44, 1954.

[11] W. K. Baker and E. von Halle, *Science*, **119**, 46–48, 1954.

[12] N. H. Giles, *Genetics*, **28**, 398–418, 1943.

[13] W. K. Baker, G. H. Mickey, and A. F. Yanders, Operation Upshot/Knothole, Project 23.6, January, 1954, pp. 21–23.

*EVIDENCE OF COMPLEX LOCI IN SALMONELLA**

By M. Demerec, I. Blomstrand, and Z. E. Demerec

CARNEGIE INSTITUTION OF WASHINGTON, DEPARTMENT OF GENETICS, COLD SPRING HARBOR, NEW YORK

Communicated April 25, 1955.

The first step in our program to study spontaneous and induced mutability in individual genes of *Salmonella typhimurium* was the accumulation of more than 250 mutants defective in the synthesis of various amino acids or purines or in the fermentation of galactose. All these mutants arose independently of one another. The first few were found in cultures treated with ultraviolet rays, whereas the great majority were obtained from untreated cultures in which they originated as spontaneous mutants. Our collection now includes about 50 strains deficient for cystine, 35 deficient for histidine, 39 for methionine, 22 for proline, 35 for serine, 10 for tryptophan, 13 for adenine, and 12 for adenine and thiamine, as well as 46 strains that are unable to ferment galactose.

As a second step, we undertook to determine which mutants showing the same phenotype were allelic to one another, using a simple and reliable method based on the work of Zinder and Lederberg[1] on transduction. These workers discovered that phage raised on wild-type bacteria is able to induce changes to wild type in a small fraction (one in 10^5 to 10^3) of mutant bacteria infected with it but that phage raised on mutant bacteria cannot transduce another population of the same strain of mutant bacteria. All the evidence we have accumulated so far supports the hypothesis proposed by Zinder and Lederberg that a phage particle is able to transfer genetic specificity—presumably by carrying a fragment of a chromosome from the bacterium on which it was raised (donor bacterium) into the bacterium which it infects (recipient bacterium). During subsequent division of the recipient bacterium, this chromosome fragment is incorporated into a chromosome of its offspring, and thus certain genes of the recipient bacterium are replaced by homologous genes of the donor bacterium. With such a mechanism in operation, it is evident that the results of transduction can be detected only if the donor and the recipient bacteria differ in genetic constitution, and therefore transduction experiments can be used to determine whether or not two similar mutants are genetically identical. For example, transduction between two cystineless strains would not be detectable if the deficiencies they carried were due to modifications at identical sites on their chromosomes, but it would be detectable if these modifications were not identical. Therefore, the method of testing similar bacterial mutants for identity consists in growing phage on each of the mutant strains and then using this phage in transduction experiments to infect each of the strains.

Grouping among Phenotypically Similar Mutants.—We began transduction tests with a collection of cystineless strains (*cys-1*, *cys-2*, etc.). No transduction was observed between *cys-20* and *cys-1*, *-3*, *-5*, *-13*, *-21*, or *-22*, indicating that the last six are allelic to *cys-20*. Also no transduction was discernible between *cys-3* and *cys-5*, indicating allelism between them. Transduction was evident between all the other members of the group *cys-1*, *-3*, *-5*, *-13*, *-20*, *-21*, and *-22* (designated group A); but the number of transductions was considerably smaller than the

number observed either when wild-type phage or phage grown on other *cys* strains was used with bacteria of group A or when other *cys* bacteria were used with group A phage (see M. Demerec[2] [Table 4]). A similar relationship was found among *cys-10*, *-12*, *-14*, *-15*, *-16*, *-18*, and *-24* (group B), except that in this group a few transductions were obtained in each combination.

Continuing these analyses of transduction between cystine-requiring mutants of independent origin, we were able to place 42 such mutants in three groups, namely, group A, with 8 members; group B, with 20 members; and group C, with 14 members. Similarly, 23 serine deficiencies fell into two groups, 22 proline into two groups, 10 tryptophan into four groups, 23 methionine into five groups, 34 histidine into four groups, 12 adenine into five groups, and 12 adenine-thiamine into four groups. Such grouping has been observed in every case when this type of analysis has been applied to several mutants independent in origin but similar as to phenotype.

Further studies of these groups have revealed that members of the same group may have in common certain properties that are absent, or present to a different degree, in the members of other groups. For example, the cystine deficiency in all but one of the members of group A could be partially satisfied by methionine, and to a somewhat lesser degree by homocysteine or cystathionine, whereas this was the case with only about 50 per cent of the members of group C and only one out of the 20 members of group B.

A more striking example of correspondence between the grouping arrived at by transduction experiments and grouping in terms of some other characteristic was found in studies of galactose-negative mutants carried on by Z. E. Demerec. Forty-six independently originating mutants could be separated into eight classes on the basis of three criteria: appearance of colonies on EMB-plus-galactose medium; amount of residual growth on synthetic medium with galactose as a carbon source; and degree of instability, as shown by the formation of sectors or papillae on colonies grown on EMB medium and by frequency of reversions to wild type on minimal medium. Only two of these classes (the second and third) were suitable for transduction experiments, because the strains belonging to the other six showed such a high degree of residual growth on our minimal medium that the usual techniques for detecting transduction could not be applied. Tests with the 8 mutants of class 2 showed that all belonged to one transduction group and that none of the 5 mutants of class 3 was in that group.[3]

The best evidence of coincidence between groups based on transduction and groups based on some other property, however, is now being accumulated in biochemical studies, carried on by several members of our laboratory, of blocks in the chains of reaction leading to synthesis of the compounds required by phenotypically similar auxotrophs. An investigation of our 10 tryptophan-requiring mutants, made here by Dr. Sydney Brenner, of the department of physiology, Medical School, University of the Witwatersrand, Johannesburg, South Africa, showed agreement between the grouping of these mutants based on transduction tests (groups A, B, C, and D) and that indicated by biochemical analysis. He found that the 1 mutant in group A failed to synthesize anthranilic acid; the 2 in group B were unable to convert anthranilic acid to an as yet unidentified intermediate, called compound "B"; the 1 group-C mutant failed to convert compound B to indole;

and each of the 6 group-D mutants was blocked in the conversion of indole to tryptophan. Dr. Philip E. Hartman's analysis of the 34 histidine-requiring strains shows that the 9 members of group A do not accumulate imidazoles; the 7 group-B mutants accumulate imidazole glycerol; the 6 group-C mutants accumulate substances tentatively identified as imidazole lactic acid, imidazole acetol, and an unknown imidazole; and 11 of the 12 group-D mutants accumulate histidinol, imidazole acetole, imidazole lactic acid, and small amounts of imidazole glycerol. Thus, so far, there is good agreement, in our tryptophanless and histidineless mutant strains, between transduction grouping and grouping in accordance with nutritional requirement or accumulation of certain substances. Biochemical studies of blocks in cystineless strains, carried on by Z. E. Demerec, have indicated a similar pattern with regard to these mutants also. Tests with several sulfur compounds showed that the cystine requirement can be satisfied in members of group A by Na_2SO_3 and in members of group C by $Na_2S_2O_3$ and Na_2S but that group B-deficiencies cannot be satisfied by any of these compounds.

The results just described show that phenotypically similar mutants can be separated by transduction tests into well-defined groups. Between members of the same group, transduction to wild type either does not take place or—in the great majority of instances—is significantly less frequent than it is between these members and mutants belonging to another group, or wild-type bacteria. They also show that among these auxotrophs the grouping based on transduction tests coincides with that obtained by biochemical methods which investigate blocks in the chains of reaction leading to synthesis of the compounds required by the auxotrophs.

Discussion.—When we began the first transduction tests to determine allelic relationships among 25 cystine-requiring mutants, we were greatly surprised by the results, which showed that transduction to wild type occurred in all but a very few combinations. It did not seem reasonable to suppose that almost twenty gene loci, each playing a major role in the synthesis of cystine, could exist in *Salmonella*. The key to solution of this puzzle was the observation that *cys-20* produced no transductions with *cys-1*, *-3*, *-5*, *-13*, *-21*, or *-22*, indicating allelism between *cys-20* and these other cystineless mutants, but that transductions to wild type were obtained in all combinations between *cys-1*, *-3*, *-13*, *-21*, and *-22*, suggesting no allelism between these mutants and thus contradicting the previous indication. The same experiments, however, revealed that the numbers of transductions between members of this group were considerably smaller than the numbers obtained when any of them was tested either with wild type or with other cystineless mutants of the collection. Subsequent experiments, made with cystineless and other auxotrophic mutants, demonstrated the general presence of such groups among similar mutants and showed that the same groupings could be established by studies of transduction and by studies of biochemical blocks. These results disclosed a considerably higher degree of similarity between mutants belonging to a certain group than between mutants belonging to different groups, and they also reduced the number of major classes within any one mutant phenotype to a level that could reasonably be accounted for by the assumption that all mutants belonging to the same class are nonidentical alleles of the same gene locus. If this is so, then a gene locus extends over a section of a chromosome; changes occurring in different regions of this sec-

tion give rise to different alleles; and the transduction observed between alleles is due to incorporation into the transduced chromosome of a normal, unchanged region brought by phage from the donor to the recipient bacterium.

It is reasonable to assume that the incorporation of a chromosome fragment in transduction is accomplished during bacterial division, at the time of duplication of the bacterial chromosome. It could be brought about either through direct incorporation of the fragment into the newly formed chromosome (Fig. 1, *A*) or by the fragment's undergoing duplication simultaneously with the bacterial chromosome, replicas of the fragment being incorporated into the new chromosome by a process similar to the one proposed by Belling[4] in his explanation of the mechanism of crossing-over (Fig. 1, *B*). In the latter case, transduction would require an even number of crossovers between the chromosomal strand derived from the bacterial chromosome and the strand derived from the fragment brought in by the phage. Our experimental evidence favors the last-mentioned possibility.

For consideration of our present problem, it is not essential to know which of these two mechanisms operates in transduction, and therefore discussion of that phase of the problem will be reserved for another paper. We are working on the assumption that the chromosome fragment brought in by phage may exceed the

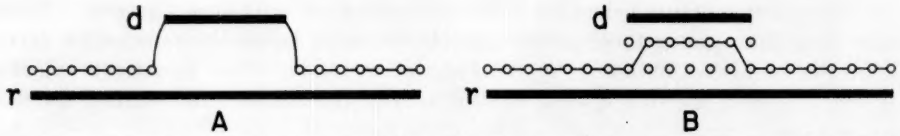


FIG. 1.—Diagram illustrating two possible mechanisms by which a fragment of genetic material brought by phage into a recipient bacterium might be incorporated into the duplicating bacterial chromosome. *A*, direct incorporation; *B*, incorporation by double crossing-over (*r:* chromosome of recipient bacterium; *d:* fragment brought from donor bacterium).

length of a gene locus and that the ends of the incorporated fragment need not correspond with the ends of a locus. As already pointed out, we assume that in the case of two nonidentical alleles the changes (mutations) responsible for the phenotypic difference between wild type and mutants have occurred in different regions of the gene locus. Then transduction to wild type will take place only when the new chromosome derived from chromosomes of the donor and recipient bacteria is composed of segments of the involved locus which do not have mutated regions. It is evident that the frequency with which such combinations arise will be positively correlated with the distance between the sites within a gene locus where changes responsible for alleles have occurred. Thus the frequency of transductions between nonidentical alleles—different members of the same group—might serve as a means of determining the linear order of alleles within a locus and the distances between them. Unfortunately, the efficiency of transduction is affected by several factors that we are not yet able to control, so that quantitative data obtained in different experiments are not always comparable and consequently the values for recombination between nonidentical alleles give only an approximate indication of the distances between these alleles and their order of arrangement within a locus. For example, the data given in Table 1 show that the number of transductions among *cys-14*, *-16*, *-18*, and *-24* is considerably smaller than between any of these

and *cys-10*, *-12*, or *-15*, indicating that *cys-14*, *-16*, *-18*, and *-24* are located close together and some distance from *cys-10*, *-12*, and *-15*.

Recently we have found linkage between several loci, which makes it possible to use three or four easily distinguishable markers in individual experiments and opens up a more reliable way of determining the order of alleles at these loci and the distances between them.

Among the 28 cystineless mutants of groups A and B, we have found only 6 cases in which transductions to wild type did not occur between certain members of a group: there were none between *cys-20* and any of the other 7 members of the *cysA* group or between *cys-3* and *cys-5* or between *cys-2*, *-6*, *-30*, and *-39*. Similarly, Hartman, in analyzing the 34 histidineless mutants, found 9 cases in which transduction was not observed. T. Yura, analyzing 12 adenineless mutants, found that 2 members of the *adA* group and 2 out of 8 members of the *adB* group gave no transductions with each other; and in an analysis of 12 strains requiring adenine plus thiamine he found that 3 of the 8 members of group A and both members of group D failed to show transductions with each other. With regard to the 10 tryptophanless strains, we found that transduction was not observed between 2 of the 6 members of group *tryD*. In the great majority of cases, however, mutants belonging to any particular group produce wild-type transductions with the other members of that group, showing that they are genetically different. This suggests the existence of many sites within a gene locus where changes responsible for the appearance of nonidentical alleles may occur. In cases in which a mutant fails to produce transductions with several nonidentical alleles of the same locus, we may be dealing with some chromosomal aberration like a deficiency or an inversion rather than with a more common change involving a single site.

TABLE 1

NUMBER OF TRANSDUCTIONS OBSERVED IN EXPERIMENTS IN WHICH ABOUT 3×10^8 BACTERIA AND 2.3×10^9 PHAGE PARTICLES WERE USED IN EACH COMBINATION

	Phage Raised on								
Bacteria	*cys-10*	*cys-12*	*cys-14*	*cys-15*	*cys-16*	*cys-18*	*cys-24*	*Wild*	*cys-11*
cys-10	0	47	42	65	26	29	29	1,115	1,157
cys-12	28	0	43	83	33	36	25	2,104	2,351
cys-14	29	38	0	33	2	4	6	1,792	2,220
cys-15	76	95	107	0	54	88	61	954	1,019
cys-16	15	19	2	34	0	2	3	979	1,223
cys-18	5	20	1	23	3	0	3	771	856
cys-24	25	47	14	84	10	8	0	1,617	528
cys-23	1,879	791	2,296	1,399	951	860	1,616	1,592	742

Summary.—By the use of transduction tests, phenotypically similar auxotrophic mutants of *Salmonella* can be separated into well-defined groups, and this grouping corresponds with the grouping arrived at by studies of their biochemical blocks. It is assumed that each such group represents a gene locus and that different members of a group are "nonidentical alleles (pseudo-alleles)," which have originated through mutations occurring at different sites in a gene locus.

* Supported in part by a grant-in-aid from the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council.

[1] N. D. Zinder and J. Lederberg, "Genetic Exchange in Salmonella," *J. Bacteriol.*, **64**, 679–699, 1952.

[2] M. Demerec, "What Is a Gene?—Twenty Years Later," *Am. Naturalist*, **89**, 5–20, 1955.

[3] M. Demerec, H. Moser, J. Hemmerly, I. Blomstrand, Z. E. Demerec, P. L. FitzGerald, S. W. Glover, J. F. Hanson, F. J. Nielsen, and T. Yura, "Bacterial Genetics—I," *Carnegie Inst. Wash. Year Book*, **53**, 225–241, 1954.

[4] J. Belling, "Crossing Over and Gene Rearrangement in Flowering Plan ts," *Genetics*, **18**, 388–413, 1933.

---

*GENETICALLY CONDITIONED METABOLIC ANTAGONISM AND SUPPRESSOR ACTION IN NEUROSPORA**

BY C. O. DOUDNEY† AND R. P. WAGNER

GENETICS LABORATORY, DEPARTMENT OF ZOÖLOGY, UNIVERSITY OF TEXAS, AUSTIN, TEXAS

Communicated by J. T. Patterson, March 31, 1955

A mutant strain of *Neurospora crassa*, T77, which grows at a rate equivalent to that of the wild-type strains at 35° C. but is markedly inhibited in growth by the amino acid *l*-threonine has been previously described.[1, 2] By inhibition analysis it was demonstrated that the threonine inhibition of this strain could be explained as a competitive antagonism of threonine to homocysteine metabolism which results in the interruption of the biosynthesis of methionine, the thiazole portion of thiamine, and, at higher threonine concentrations, adenine and serine.[2] Wagner and Bergquist[3] have shown that this strain accumulates α-ketoisovaleric, α-keto-β-methylvaleric, and pyruvic acid in its growth medium when growing in the presence of threonine, but not in its absence. The accumulation of these keto acids is accompanied by a rapid decrease in the threonine concentration during a period of time in which there is only a small increase in the weight of the mycelium. Wild-type strains do not accumulate keto acids, except for a very small amount of pyruvate when growing in the presence of threonine. The data suggest that the inhibitory effect of threonine on homocysteine metabolism in this strain is in some way related to the rapid disappearance of threonine and the consequent accumulation of keto acids which can be presumed to be in part, at least, products of threonine metabolism,[3] though the biochemical mechanism involved cannot be resolved on the basis of present evidence.

The early analysis of this strain led to the suggestion[1] that the inability of an organism to mediate a chemical synthesis because of a genetic block may not necessarily be the result in every case of the absence or inactivation of an enzyme required for that synthesis but may be caused by some sort of inhibition at the metabolic level resulting from the mutation. The studies described here were undertaken with the objective of testing this hypothesis by the use of strain T77. If the aberrant metabolic mechanism responsible for the complex nutritional requirements of T77 in the presence of threonine could be shown to be operative in the absence of threonine under altered environmental conditions, thus causing a requirement for exogenous sources of metabolites normally synthesized by the wild-type strain under similar environmental conditions, then this hypothesis would receive some support. The results described below demonstrate that T77 requires either the thiazole portion of thiamine or both homocysteine and threonine for growth at lower temperatures of incubation, indicating the operation of the inhibitory mechanism in growth failures at lower temperatures.

If metabolic antagonisms or inhibitions are the basis of some genetic blocks to biosynthesis, then it should be possible to suppress these blocks by any method which leads to a reduction in the synthesis of the inhibitory metabolite. Thus a loss of ability to form the inhibitory metabolite by introduction of a mutant gene would lead to suppression of the original requirement. Wagner and Haddox[4] have described a case of suppression which seems to be of this nature in a double mutant strain containing gene mutations which in separate strains produce requirements, respectively, for pantothenic acid and tyrosine. In combination, only the tyrosine requirement remained, the pantothenic acid requirement being suppressed. We have tested this hypothesis experimentally in the case of the inhibitory mechanism operative in mutant strain T77 and have shown that introduction into the strain of a block to threonine biosynthesis suppresses the growth requirement for thiazole at lower temperatures of incubation and the growth inhibition by threonine at higher temperatures of incubation.

Methods.—The techniques utilized in these studies are those previously described.[1, 2] In the construction of the double mutant strains, crosses were made on Westergaard medium supplemented with *dl*-threonine (1×10^{-4} *M*). The ascospores obtained from these crosses were dissected from the asci in order. The resultant strains were tested for growth characteristics, and the double mutant strains were selected on the basis of the genetic results as well as on the basis of their growth requirements. That the strains selected actually possessed the mutant character furnished by each parent strain was determined by crossing the strains to a wild-type strain from which both expected mutant strains were regained.

Incubation Temperature Studies.—Figure 1 presents the growth response of T77 in 72 hours at various temperatures of incubation with and without the thiazole portion of thiamine added as a nutritional supplement. At 20° C. this strain does not grow to any appreciable extent (less than 0.5 mg.). However, the addition of small amounts of thiazole raises the growth response of the strain to the level of wild-type strains. In addition to thiazole, the combination of threonine and homocysteine will promote growth equivalent to that of wild-type strains (Fig. 2). Either of these amino acids added alone is only slightly stimulatory. Thus T77 may be characterized as a temperature-sensitive strain having a requirement for either thiazole or both threonine and homocysteine at 20° C. but growing at a rate equivalent to wild-type strains at 30° C. or higher. This is in addition to its above-mentioned sensitivity to threonine inhibition at 35° C.

Threonine-Homocysteine Antagonism.—T77 grows at a reduced rate at 25° C. when compared to the wild-type strain (Fig. 3). The addition of either thiazole or both homocysteine and threonine will restore growth to the wild-type level. At this incubation temperature, either homocysteine or threonine, when added alone in nutritional concentrations (1×10^{-4} *M*), is somewhat stimulatory to growth. However, higher concentrations of either of these amino acids become inhibitory at this temperature. An equimolar supplement of these amino acids at any concentration is not inhibitory. Thus the presence of either amino acid in the medium produces a competitive requirement for the other amino acid. Methionine may be substituted for homocysteine in effecting this threonine competitive growth inhibition. However, methionine reverses threonine inhibition noncompetitively.[1] It would thus appear that, while methionine may act (presumably through transformation to homocysteine) in the inhibitory mechanism which inter-

feres with threonine metabolism, it is past the point of interference by threonine with homocysteine metabolism.

It is interesting to note that at 30° C. this strain is not inhibited by high concentrations of either threonine or homocysteine. At 35° C., however, as has previously been shown,[1] the strain is markedly sensitive to threonine in the medium. At this temperature homocysteine is even less inhibitory than it is to wild-type strains.[5]

Various antagonisms involving methionine, threonine, and related four-carbon amino acids in the growth of *Neurospora* have been demonstrated. Emerson and Cushing[6] and Emerson[7] reported a mutation, *sfo*, in *Neurospora* which causes a requirement for sulfanilamide at 35° C. and pH below 6. Zalokar[8] postulated that the mutation responsible for the sulfanilamide requirement controls a deleterious reaction which, using a precursor of methionine as a substrate, is catalyzed

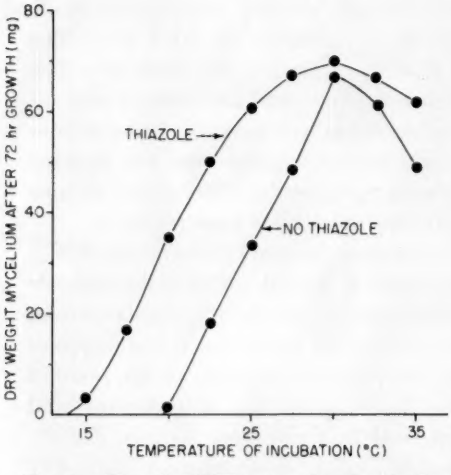


FIG. 1.—Growth of T77*a* at various incubation temperatures in the presence and in the absence of the thiazole portion of thiamine (0.5 μg/25 ml).

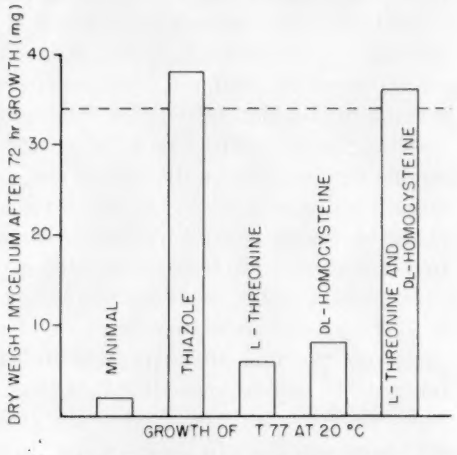


FIG. 2.—The influence of thiazole, homocysteine, and threonine on growth of T77*a* at 20° C. Concentrations: thiazole, 0.5 μg/ml; *l*-threonine, 1×10^{-4} *M*; *dl*-homocysteine, 1×10^{-4} *M*. The broken line indicates the level of growth attained by wild type under these conditions.

by *p*-aminobenzoic acid and interferes with the normal utilization of threonine. Doudney and Wagner[1] postulated a similar reaction as a basis for the homocysteine requirement of the threonine-inhibited strain of *Neurospora*. However, up to the present time no adequate biochemical basis for these examples of threonine-homocysteine antagonism has been presented. It is difficult to explain the mutually antagonistic effect of these amino acids at lower temperatures on the basis of competition for an enzyme unless one assumes that the enzyme has a function common to both metabolic sequences. More plausible as a basis for this mutual antagonism is a competition of both the threonine and homocysteine metabolic systems for a metabolite common to both. However, concrete evidence for a metabolite common to both these systems is not available. It is evident that more work at the biochemical level will be necessary before the basis for the double amino acid requirements can be fully evaluated.

These results strongly suggest, however, an antagonism of endogenously produced threonine to homocysteine metabolism as a basis for the thiazole requirement of the strain observed at 20° C. This type of metabolic antagonism may be highly significant as one basis for genetic blocks to biosynthesis.

Gene Interactions.—In order to determine the effect of genetic interruption of threonine biosynthesis on the threonine inhibition process, double mutant strains possessing gene mutation responsible for threonine inhibition and a gene mutation causing a block in threonine biosynthesis were constructed. Strain T77 was crossed to two threonine-requiring strains.[9] One strain (35423)[10] has a requirement specifically for threonine and is presumably blocked in a later stage of threonine biosynthesis. A competitive antagonism between methionine and threonine has been observed in this strain. The other strain (44104)[10] can make use of either threonine or α-aminobutyric acid and, to a limited extent, of isoleucine. This strain is presumably blocked at an initial step to threonine biosynthesis from a four-carbon metabolite related to homoserine.

Figure 4 presents the growth responses to threonine at 35° C. obtained from double mutant strains possessing these mutant genes in combination with the T77 gene as compared to the parent strains. The response of 44104, T77 is equivalent to that of the 44104 parent strain. At 10^{-3} M, the concentration of maximum response of the double mutant strain, *l*-threonine is maximally inhibitory to T77.

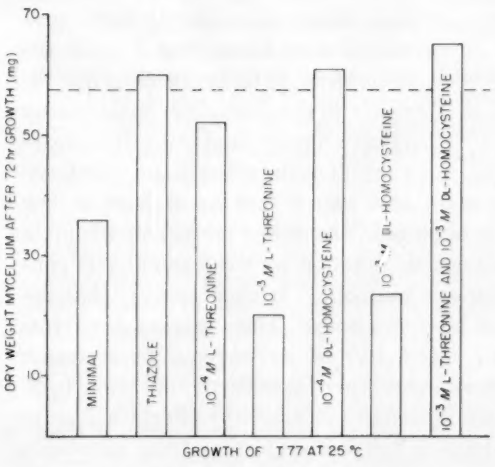


FIG. 3.—The influence of thiazole (0.5 μg/25 ml), homocysteine, and threonine on growth of T77*a* at 25° C. The broken line indicates the level of growth attained by wild type under these conditions.

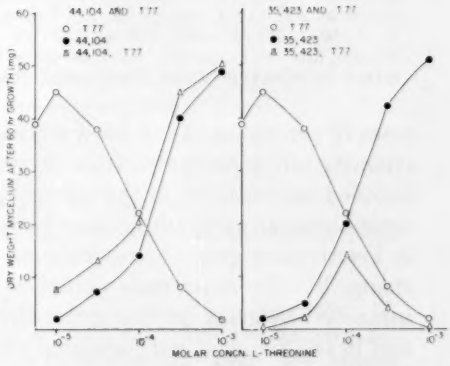


FIG. 4.—A comparison of growth responses to threonine at 35° C. of T77*a*, two threonine-requiring strains (35423, 44104), and the double mutant strains obtained from crosses between T77 and the threonine-requiring strains.

As Figure 5 shows, this gene mutation suppresses the growth requirement produced by the T77 gene mutation at 20° C., since the double mutant strain requires threonine but not the thiazole moiety of thiamine at this incubation temperature. It may be postulated that the suppression is due to the failure of the double mutant strain to synthesize threonine, since an increase of threonine concentration beyond that used in Figure 5 will re-create the growth requirement for thiazole.

The growth responses of the 35423, T77 strain to threonine are quite different. The response of this strain is antagonized by a marked inhibitory effect of the amino acid (Fig. 4). The basis for the failure of this genetic block to threonine biosynthesis to suppress threonine inhibition is not apparent. One possibility is that 35423 is not a simple genetic block based on an absent or inactive enzyme but is similar to that postulated for T77, with the threonine requirement due to an interaction between homocysteine and threonine. This possibility is supported by Teas'[9] observation that methionine or homocysteine antagonizes threonine utilization in this strain. It would be expected, if this were the case, that the double mutant strain having a double genetic basis for homocysteine-threonine antagonism would be subject to threonine inhibition.

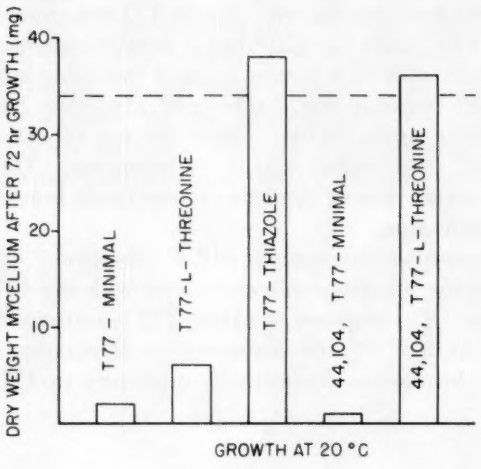


FIG. 5.—The thiazole and threonine requirements of T77*a*, 44104 and the double mutant strain at 20° C. Concentrations: *l*-threonine, 1×10^{-4} *M*; thiazole, 0.5 μg/100 ml. The broken line indicates the level of growth obtained by wild type under these conditions.

Discussion.—While accumulating evidence indicates a rather direct relationship of gene action to enzyme formation, only in a few cases has clear-cut evidence been presented to indicate that a given genetic block to biosynthesis is due directly to the absence or inactivation of an enzyme mediating that reaction. On the other hand, evidence also exists that in at least a few cases of genetic blocks to biosynthesis a functional enzyme is present in adequate quantity but is prevented from carrying out its function by some genetically conditioned modification in the intracellular environment. It thus appears that genetic blocks to biosynthesis may have at least two bases. One basis is a direct loss or inactivation (e.g., a modification of specificity) of an enzyme by a genetic change.[11] The other basis underlies those cases where genetic modification (presumably producing an enzymatic change in amount or activity) effects a disruption in the dynamic and correlated processes of metabolism leading to an imbalance in cellular biosynthetic economy. As a result of this imbalance, an inhibitory metabolite may accumulate or an essential metabolite may be utilized more rapidly in one reaction than it is supplied for others.[4]

An aberrant system akin to that described above seems to operate in the case of the growth requirement of T77 at 20° C. The data support the following hypothesis. By a genetically conditioned inhibitory mechanism as yet not clearly elucidated, endogenously formed threonine interferes with homocysteine utilization. A decrease in rate of this system occurs, presumably setting up a severe competition among all those biosyntheses dependent on homocysteine metabolism. Some are able to proceed (e.g., methionine formation), but their successful competition for the limited supply prevents adequate synthesis of the thiazole portion

of thiamine. Consequently, growth failure based on a thiazole requirement results. Since the synthesis of threonine is responsible for the inhibition, a reduction of threonine biosynthesis by genetic means results in suppression of the growth requirement.

The possibility exists that most if not all cases of suppression by one genetic character of the expression of another occurs through mechanisms operative at the metabolic or enzymatic level. Thus consideration of the several observations of genetic suppressors to nutritional requirements in *Neurospora*[12] lends support to the view that metabolic antagonisms may be an important basis for genetic blocks to biosynthesis. It is now desirable to assess the relative occurrence of these imbalances in the production of various genetic blocks and to investigate the possible significance of genetically conditioned metabolic interactions in developmental and evolutionary phenomena.

Summary.—A mutant strain of *Neurospora crassa* which grows like wild-type strains at 35° C. but which is markedly inhibited by the amino acid *l*-threonine requires the thiazole moiety of thiamine at lower temperatures of incubation for normal growth. The basis of the thiazole requirement may be a competitive antagonism between threonine and homocysteine. The requirement for thiazole is suppressed by introduction of a genetic block to threonine biosynthesis. The relation of these observations to possible mechanisms of genetic blocks to biosynthesis and suppressor action is discussed.

* Aided by funds from the Office of Naval Research, United States Navy Department, administered by the University of Texas under contract No. Nonr 859(00).

† Present address: Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

[1] C. O. Doudney and R. P. Wagner, these PROCEEDINGS, **38**, 196, 1952.

[2] C. O. Doudney and R. P. Wagner, these PROCEEDINGS, **39**, 1043, 1953.

[3] R. P. Wagner and A. Bergquist, *J. Biol. Chem.*, 1955 (in press).

[4] R. P. Wagner and C. H. Haddox, *Am. Naturalist*, **85**, 319, 1951.

[5] C. O. Doudney and R. P. Wagner, *Arch. Biochem. and Biophys.*, 1955 (in press).

[6] S. Emerson and J. E. Cushing, *Federation Proc.*, **5**, 379, 1946.

[7] S. Emerson, *J. Bacteriol.*, **54**, 195, 1947.

[8] M. Zalokar, *J. Bacteriol.*, **60**, 191, 1950.

[9] H. J. Teas, N. H. Horowitz, and M. Fling, *J. Biol. Chem.*, **172**, 651, 1948.

[10] Obtained through the courtesy of Dr. N. H. Horowitz, of the California Institute of Technology.

[11] G. W. Beadle, *Chem. Revs.*, **37**, 15, 1945.

[12] N. H. Giles, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 238, 1951; N. H. Giles and C. W. H. Partridge, these PROCEEDINGS, **39**, 479, 1953; J. Lein and P. S. Lein, these PROCEEDINGS, **38**, 44, 1952; M. B. Mitchell and H. K. Mitchell, these PROCEEDINGS, **38**, 205, 1952; B. S. Strauss and S. Pierog, *J. Gen. Microbiol.*, **10**, 221, 1954.

*NUCLEO-CYTOPLASMIC RELATIONS IN SALIVARY-GLAND CELLS OF DROSOPHILA**

By Helen Gay

CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, NEW YORK, AND UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PENNSYLVANIA†

Communicated by Berwind P. Kaufmann, April 4, 1955

The importance of the nucleus in the control of cellular functions has been demonstrated by studies in many fields of experimental biology.[1] The results have indicated that sustained physiological activity of the cell and the expression of its genetic potentialities are dependent on the transfer of materials between nucleus and cytosome. Theoretical considerations of nucleo-cytoplasmic relations have engaged the attention of biologists since de Vries advanced the hypothesis of intracellular pangenesis in 1889, but the supporting cytological evidence has been extremely meager, being for the most part a series of observations of cytoplasmic particles that assumedly were of nuclear origin.[2] In some types of cells, such as microsporocytes of various plants,[3] these cytoplasmic particles were Feulgen-positive, indicating that they contained deoxyribonucleic acid; but it was not always possible to determine whether they had been ejected from the nucleus in the normal course of cellular activity or as a result of degenerative changes. In other types of cells, such as the oöcytes of amphibians and fishes,[4] the cytoplasmic particles were identified as nucleoli; and extrusion of nucleolar material through the nuclear membrane was reported to be induced by treatment of the living amphibian oöcyte with dilute acid.[5] Such findings indicated that particles of large size might pass through the nuclear membrane, although the mechanism responsible for their transport in the normal metabolizing cell was not determined. In the course of an electron microscope study of the salivary-gland cells of *Drosophila melanogaster*, structural configurations were observed that suggested a mechanism for exchange of materials between nucleus and cytosome.[6] The electron micrographs showed that outpocketings of the nuclear membrane were intimately associated with specific regions of the chromosomes. A survey of the experimental evidence is presented here.

Materials and Methods.—Salivary glands of well-fed third-instar larvae of *D. melanogaster* (Sw-b 6) were excised and fixed in ice-cold 1 per cent osmium tetroxide, isotonic with insect blood,[7] and buffered at pH 6.5.[8] The glands were imbedded in *n*-butyl methacrylate[9] and sectioned with a Porter-Blum ultramicrotome. Serial sections prepared by the method of Gay and Anderson[10] were examined in a Philips electron microscope without removal of the imbedding medium.

Results.—Electron micrographs of these ultrathin sections of salivary-gland cells showed well-preserved structural elements in the cytosome, including spherical secretion granules, mitochondria, and abundant endoplasmic reticulum in the form of narrow membranous sacs aligned parallel to each other and to the contours of the nucleus (Fig. 1). The nucleus contained fibrous chromosomes, whose outlines were readily discernible, although no limiting membranes were apparent. The nuclear membrane was composed of two layers, separated by a distance of about 150 A; they were most clearly defined in cross-sections that were normal to the

nucleus (perpendicular to the tangent at the level of the section). The contour of the nucleus was undulous, and in some places the membrane protruded into the cytosome in the form of outpocketings or blebs (the arrows of Fig. 1).

It seemed necessary to establish with certainty that these outpocketings were not attributable to fixation distortion (even though the fine structural details characteristic of good fixation were apparent in cytoplasmic organelles[11]), and therefore a lengthy series of modifications of technical procedures was undertaken. They included variations in the salt concentration and pH of the fixative, in the time and temperature of fixation and subsequent treatment, and in the nature of the agents used in the postfixation processes. None of these modifications was effective in eliminating the uneven border of the nucleus and the outpocketings of the membrane. It should also be noted that undulatory nuclear contours have been reported for other types of secretory cells.[12] On the basis of these considerations, it seems reasonable to conclude that outpocketings of the nuclear membrane are characteristic structural features of the salivary-gland cells of third-instar larvae of *D. melanogaster*.

The outpocketings were not uncommon in this material, and single ultrathin sections sometimes showed as many as five or six. More significant than the number, however, was the fact that the chromosomal materials adjacent to the blebs appeared different from those in other parts of the nucleus. The strands were more highly electron-scattering and were often associated with a type of material that was not discernible in the body of the chromosome. The chromatin strands were in contact with the nuclear membrane at the region of blebbing and in some cases were observed to extend into the outpocketings (Figs. 2, 3, and 4).

Examination of serial sections in which an outpocketing could be traced through its entirety invariably revealed connections with the salivary-gland chromosomes. In some cases the connections were to single bands in intercalary regions and in others apparently to terminal chromatin. Occasionally the intercalary band was located in a "reverse repeat" (Figs. 6*a*–6*c*). No positive identification was possible in the ultrathin sections with respect to the specific subdivisions of chromosomes involved, but it should be kept in mind that some reverse repeats have been shown to contain intercalary heterochromatic regions.[13]

Discussion.—Observations on static systems such as are represented in sections of fixed material have limitations with respect to the interpretation of functional relationships, but the intimate association between chromosomal materials and membrane outpocketings revealed in these electron micrographs strongly suggests a mechanism for transfer of chromosomal products to the cytosome.

In evaluating this interpretation, the possibility must be considered that the outpocketings are merely transitory manifestations of a generalized surface activity at the nucleus-cytosome border. This possibility seems to be negated, however, by the observation that chromosomal material extends into the blebs, a fact which implies that the latter are formed at portions of the membrane that lie adjacent to specific regions of the chromosomes. It should also be kept in mind that the phenomenon reported here has been observed in only one type of cell at one stage of its development. This stage precedes by about 20 hours (at the temperature used) the time at which the larval salivary glands degenerate. The question may therefore be raised whether the nuclear blebs represent the first stages of nuclear degen-

eration. Several lines of evidence indicate that they do not. Peptidase activity is high in salivary glands at the stage of development represented in these studies and increases until about six hours after puparium formation.[14] Salivary glands of late larval and early pupal stages are active in producing a cement that serves to affix the puparium to the substrate.[15] No evidence of degeneration of nuclear or cytoplasmic structures was observed in any of the electron micrographs obtained in the present investigation. Finally, it may be noted that, if nuclear outpocketings and the associated materials are indicative of degenerative changes, the relation of the blebs to specific chromosomal bands indicates that this is not an indiscriminate phenomenon. Such a localized type of degeneration seems improbable. By all these criteria, the salivary glands of mid-third-instar larvae of *Drosophila* appear to be composed of functionally active and healthy cells; and, by the same criteria, the outpocketings of the nuclear membrane appear to be a manifestation of normal cellular activity involving participation of the chromosomes.

If the structural features described above provide a mechanism for transfer of materials from nucleus to cytosome, it seems probable that the blebs become detached from the nucleus to lie in the cytosome. Several spherules whose walls showed the double-layered condition characteristic of the nuclear membrane were observed adjacent to the nuclei in the preparations studied, but conclusive proof of their origin was not possible by examination of the electron micrographs. Studies with the light-microscope of sections of salivary glands of *Drosophila* and *Chironomus* revealed that Feulgen-positive strands were oriented toward the nuclear membrane in several regions and that small droplets of faint pink-to-colorless material occurred on the cytoplasmic side of the membrane. These findings suggest that the shape and composition of the nuclear blebs and their contents may change rapidly upon detachment. The blebs, freed from their connection with the nucleus, could conceivably discharge their contents and become flattened to produce the saclike membranes of the endoplasmic reticulum. The similarity in structure of the nuclear membrane and the elements of the endoplasmic reticulum is indeed striking, as can be seen by inspecting Figures 4 and 5. As revealed in these transverse and tangential sections, both types of membrane are double-layered and are

---

PLATE I

FIGS. 1–6.—Electron micrographs of ultrathin sections of salivary-gland cells of third-instar larvae of *Drosophila melanogaster*. The marker in the upper right-hand corner of each figure indicates 1 μ. Abbreviations: *chr*, chromosome, *nu*, nucleus; *n.m.*, nuclear membrane; *cyt.*, cytosome; *s.g.*, secretion granule; *e.r.*, endoplasmic reticulum.

FIG. 1.—Section through nucleus and adjacent cytosome, showing nuclear and cytoplasmic structure. Arrows indicate regions of outpocketings of nuclear membrane.

FIG. 2.—Outpocketing of nuclear membrane, showing dense chromosomal material confluent with chromonemata.

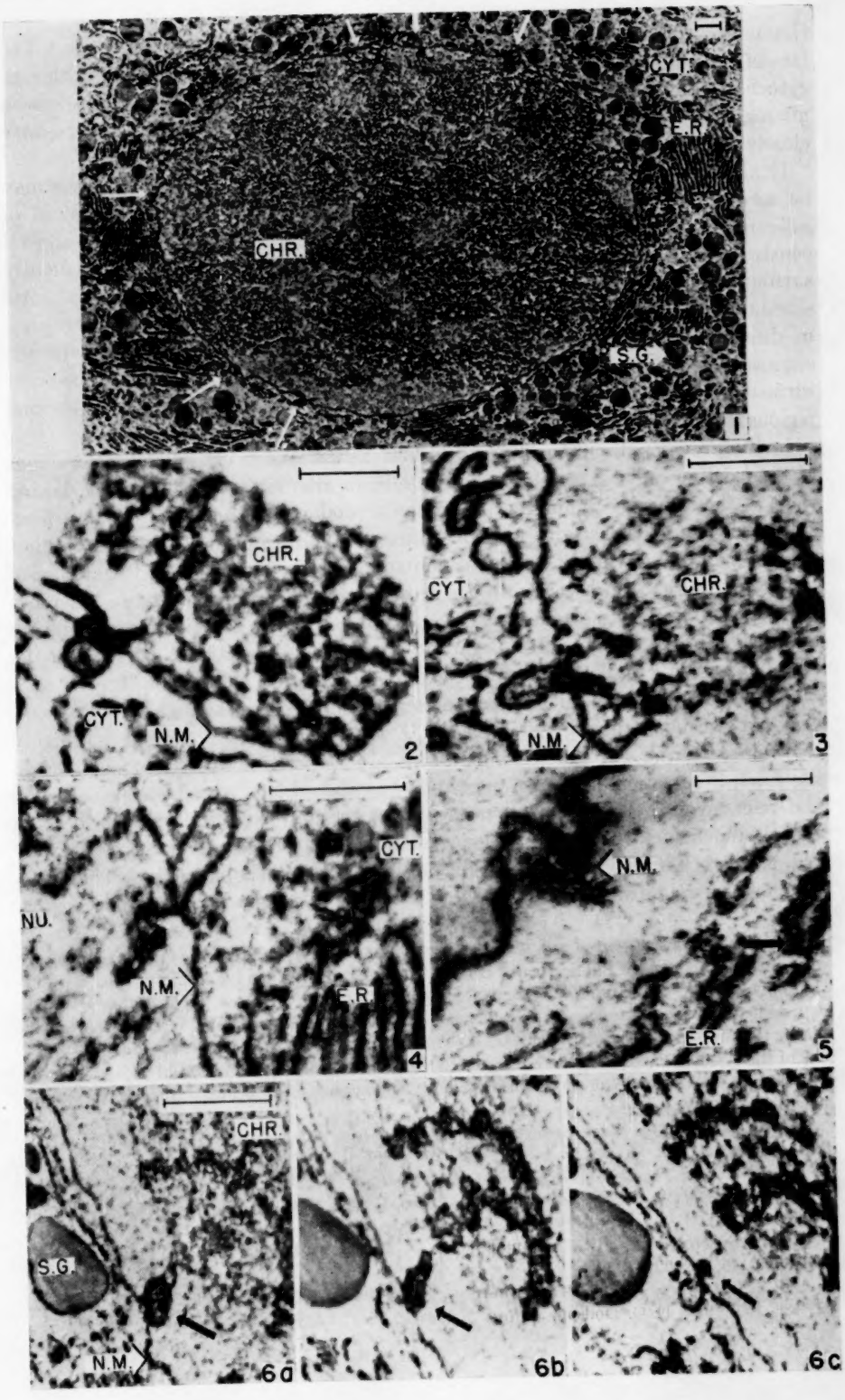
FIG. 3.—Projection from the nucleus, containing fibrous strands which connect with highly electron-scattering chromosomal material.

FIG. 4.—Outpocketing of nuclear membrane and adjacent cytoplasm. Chromatin near the bleb is electron-dense and contains spherical granules not seen in interior of nucleus. Similarity in structure of membranes of outpocketing and endoplasmic reticulum is apparent.

FIG. 5.—Tangential section through nuclear membrane and endoplasmic reticulum, showing characteristic reticulate pattern of the nuclear membrane and, at black arrow, a similar pattern of the endoplasmic reticulum.

FIGS. 6*a*–6*c*.—Three sections selected from a complete series of thirty, used to trace the connection between a single band in the chromosome and a small bleb in the nuclear membrane. Note dense connecting material at arrows.

PLATE I


CYT.
E.R.
CHR.
S.G.
1
CHR.
CYT.
N.M.
2
CYT.
CHR.
N.M.
3
NU.
CYT.
N.M.
E.R.
4
N.M.
E.R.
5
CHR.
S.G.
N.M.
6a
6b
6c

characterized by periodically spaced "pores" and associated markings. The fate of the materials included within the blebs has not been determined (although cytochemical studies directed to this end have been initiated), but, in electron micrographs of some cells, cytoplasmic structures have been identified that resemble closely the chromosomal materials extending into the outpocketings.

It has been suggested that endoplasmic reticulum in mouse pancreas cells may be associated with the production of zymogen granules.[16] If the reticulum of salivary-gland cells is associated with the formation of secretion granules—and reconstructions from serial sections have shown that large granules are frequently surrounded by the membranous sheets of the endoplasmic reticulum—an elaborate scheme could be formulated with respect to the role of specific chromosomal regions in directing metabolic processes. Such a mechanism would conform with the suggestion of Schultz[17] that the nuclear membrane, having been formed at the outer surfaces of chromosomes at telophase, may be associated with specific chromosomal regions and may therefore control specific nucleo-cytoplasmic interactions.

In the previous discussion the outpocketings have been considered as swellings of the existing nuclear membrane. There is an alternative possibility concerning their method of origin, namely, that material produced within the nucleus as a result of chromosomal activity may pass through the membrane to form a new but similar membrane when it comes in contact with the cytoplasm. Regardless of the actual mechanism involved in the production of the membranes, the existence of the blebs affords a mechanism whereby a portion of the genetic material may be transferred to the cytosome for use in the synthetic and functional activities of the cell. The genetic implications will be explored in another publication.[18]

Summary.—Electron micrographs of serial sections of functionally active salivary-gland cells of third-instar larvae of *D. melanogaster* have revealed that highly electron-scattering chromosomal materials are associated with outpocketings of the nuclear membrane. It is suggested that these blebs may become detached and released into the cytosome, where they contribute to the formation of such cytoplasmic structures as endoplasmic reticulum and secretion granules. Whether or not this is a specific mechanism for one type of cell in a particular stage of development remains to be determined.

* This study was supported in part by a research grant (RG-149) from the National Institutes of Health, United States Public Health Service, to Dr. B. P. Kaufmann, and in part by a fellowship of the Lalor Foundation for work at the University of Pennsylvania. The author wishes to express her appreciation to Drs. T. F. Anderson, B. P. Kaufmann, and C. W. Metz for their counsel during the course of this study. She also desires to thank Dr. Keith R. Porter for use of the facilities of the cytology laboratory at the Rockefeller Institute for Medical Research.

† Present address: Carnegie Institution of Washington, Cold Spring Harbor, Long Island, New York.

[1] B. H. Willier, P. A. Weiss, and V. Hamburger, *Analysis of Development* (Philadelphia: W. B. Saunders Co., 1955).

[2] E. B. Wilson, *The Cell in Development and Heredity* (New York: Macmillan Co., 1925); E. D. P. DeRobertis, W. W. Nowinski, and F. A. Saez, *General Cytology* (Philadelphia: W. B. Saunders Co., 1949).

[3] A. H. Sparrow and M. R. Hammond, *Am. J. Bot.*, **34**, 439–445, 1947; D. C. Cooper, *Am. Naturalist*, **86**, 219–229, 1952.

[4] H. S. Chaudhry, *J. Roy. Microscop. Soc.*, **71**, 87–93, 1951; M. Wittek, *Arch. biol.* (Paris), **63**, 133–197, 1952; E. O. Dodson, *J. Roy. Microscop. Soc.*, **72**, 177–178, 1953.

[5] W. R. Duryee, *Ann. N.Y. Acad. Sci.*, **50**, 920–953, 1950.
[6] H. Gay, *Anat. Record*, **120**, 728, 1954.
[7] D. Bodenstein, *Biol. Bull.*, **90**, 148–157, 1946; J. Buck, *J. Heredity*, **33**, 3–10, 1942.
[8] G. Palade, *J. Exptl. Med.*, **95**, 285–298, 1952.
[9] S. B. Newman, E. Borysko, and M. Swerdlow, *Science*, **110**, 66–68, 1949.
[10] H. Gay and T. F. Anderson, *Science*, **120**, 1071–1073, 1954.
[11] For example, as demonstrated by G. Palade, *Anat. Record*, **114**, 427–451, 1952; K. Porter, *J. Histochem. and Cytochem.*, **2**, 346–373, 1954; A. J. Dalton, *Am. J. Anat.*, **89**, 109–134, 1951; F. Sjöstrand, *Nature*, **171**, 30–32, 1953.
[12] E. B. Wilson, *The Cell in Development and Heredity* (New York: Macmillan Co., 1925); A. A. Maximow and W. Bloom, *A Textbook of Histology* (Philadelphia: W. B. Saunders Co., 1934).
[13] B. P. Kaufmann, these PROCEEDINGS, **25**, 571–577, 1939; *J. Exptl. Zool.*, **102**, 293–320, 1946; A. A. Prokofyeva-Belgoskaya and V. V. Khostova, *Compt. rend. acad. sci. U.R.S.S.*, **23**, 270–272, 1939.
[14] E. K. Patterson, M. E. Dackerman, and J. Schultz, *J. Gen. Physiol.*, **32**, 623–645, 1949.
[15] G. Fraenkel and V. J. Brookes, *Biol. Bull.*, **105**, 442–449, 1953.
[16] J. M. Weiss, *J. Exptl. Med.*, **98**, 607–618, 1953.
[17] J. Schultz, *Exptl. Cell Research*, suppl. 2, pp. 17–43, 1952.
[18] H. Gay and B. P. Kaufmann, in manuscript.

PHENOTYPIC VARIATION AND PSEUDO-ALLELISM AT THE FORKED LOCUS IN DROSOPHILA MELANOGASTER

BY M. M. GREEN

DEPARTMENT OF GENETICS, UNIVERSITY OF CALIFORNIA, DAVIS, CALIFORNIA

Communicated by Roy E. Clausen, March 22, 1955

In previous reports[1, 2] it was demonstrated that, by use of a suppressor mutant, the otherwise phenotypically inseparable vermilion (*v*) eye-color mutants of *Drosophila melanogaster* could be separated into two classes—suppressed and unsuppressed. Moreover, pseudo-allelism is indicated, since recombination was demonstrated to occur between suppressed and unsuppressed *v* mutants. The description of a suppressor of the recessive, sex-linked forked bristle (*f*) mutants[3] plus the fact that *f* mutants recur frequently suggested that the observations made for the *v* mutants be extended to the *f* mutants.

Four independent *f* mutants, spontaneous in origin, have been used. These are f^1, f^3, f^{3n}, and f^5. All are characterized phenotypically by causing a gnarling or twisting of the bristles. The f^5 mutant differs from the others in that the microchaetes as well as the macrochaetes are forked. In the tests with the sex-linked suppressor of forked (*su-f*), the following mutants (with map locations) were used: *f* (56.7); Bar eye, *B* (57.0); carnation eye color, *car* (62.5); and *su-f* (64.0). Since f^1 is suppressed by *su-f*, a stock of the genotype f^1 *B su-f* was obtained. Tests for the suppression of any independent *f* mutant, designated f^z, were made by first obtaining ♂♂ of the genotype f^z *car*. These ♂♂ were crossed to ♀♀ f^1 *B su-f*, thereby producing ♀♀ of the genotype f^1 *B* + *su-f*/f^z + *car* +. The ♂ progeny of these heterozygous ♀♀ were scored. Those ♂ progeny phenotypically B^+ car^+ must be genotypically f^z *su-f*, and their bristle phenotype was compared with

that of their f^z *car* sibs to see if the suppressor affected the forked phenotype. Among the four mutants, f^5 and f^4 are suppressed, while f^3 and f^{3n} are unsuppressed. Thus two types of *f* mutants are indicated—suppressed and unsuppressed.

Since the *f* mutants can be classified into two groups by use of *su-f*, tests were made to detect the occurrence of recombination between suppressed and unsuppressed mutants. In the initial tests attached-X ♀ ♀ were used, since, as noted previously,[2, 4] the occurrence of reciprocal crossover between pseudo-alleles in such ♀ ♀ permits the recovery of the chromosome carrying both pseudo-allelic mutants as well as the chromosome carrying both wild-type alleles together in the same ♀. Using the mutants f^1, f^{3n}, and *r* (rudimentary wing) 2.2 units to the left, and *B* 0.3 units to the right, of the forked locus as markers, attached-X ♀ ♀ of the genotype $r\ f^1\ B\ +/f^{3n}\ +\ car$ were constituted and tested. The results of these tests are listed in Table 1. It may be noted that two classes of f^+ ♀ ♀ were recovered: nonrudimentary and heterozygous *B* ($f^+\ B/+$), and rudimentary and heterozygous *B* ($r\ f^+\ B/+$).

TABLE 1

CROSSING-OVER BETWEEN *f* PSEUDO-ALLES IN ATTACHED-X ♀ ♀

EXP. NO.	GENOTYPE ATTACHED-X ♀ ♀	PHENOTYPE EXCEPTIONAL $f^+\ B/+$ ♀ ♀	PHENOTYPE EXCEPTIONAL $r\ f^+\ B/+$	TOTAL ♀ ♀
1	$r\ f^1\ B/f^{3n}\ car$	1	..	13,552
2	$r\ f^1\ B/f^{3n}\ car;\ Cy/+;\ Ubx^{130}/+$	1	..	44,872
3	$r\ f^1\ B/f^{3n}\ car;\ Cy/+;\ In(3)T,\ Mé/+$	..	1	40,033
4	$r\ f^5\ B/f^{3n}\ car;\ Cy/+;\ In(3)T,\ Mé/+$	1	..	18,969

The occurrence of the f^+ ♀ ♀ as a consequence of crossing-over between f^{3n} and f^1 can be explained by assuming that the mutants are pseudo-allelic, with f^{3n} occupying a position to the left of f^1. In Figure 1 the genotype of the parental ♀ ♀ is diagramed. In addition, the consequences of reciprocal and nonreciprocal exchanges occurring between f^{3n} and f^1 are listed. It may be noted that the phenotypically $r\ f^+\ B/+$ ♀ recovered is in agreement with expectation for a nonreciprocal exchange. That the two ♀ ♀ phenotypically $f^+\ B/+$ were derived from reciprocal exchanges was demonstrated from the results of progeny testing these ♀ ♀. The results of the progeny tests are listed in Table 2. In addition to the parental phenotypic class, two nonparental phenotypic classes were recovered as a consequence of crossing-over to the right of the *B* locus. One class comprised those ♀ ♀ homozygous *B* with an *f* phenotype designated as f^{ex}, i.e., extreme *f*, in which the microchaetes as well as the macrochaetes are forked. This represents a phenotype more extreme than that of either f^1 or f^{3n}. The second class was made up of ♀ ♀ $r\ f^+\ B^+$ in phenotype. On the basis of these progeny tests, it can be concluded that the ♀ ♀ $f^+\ B/+$ possessed one X chromosome of the genotype $r\ f^+\ B^+$ and the homologous X of the genotype $r^+\ f^{3n}\ f^1\ B$. These results fulfill expectation for a reciprocal exchange between f^{3n} and f^1.

That the chromosome designated $f^{ex}\ B$ did in fact carry both f^{3n} and f^1 was proved in the following manner. Detachments of the X chromosomes of attached-X ♀ ♀ homozygous $f^{ex}\ B$ were obtained. On testing with the *su-f*, it was observed that f^{ex} is unsuppressed. Furthermore, ♀ ♀ $r\ f^{ex}\ B/+++$ and heterozygous for the autosomal inversions *Cy* and *In(3)T, Mé* were obtained. If f^{ex} represents f^{3n} and f^1 coupled on the same X chromosome, then from these ♀ ♀ it should be possible,

by virtue of crossing-over between f^{3n} and f^{1}, to recover, among their ♂ progeny, ♂♂ $r\ f^{3n}$ and $f^{1}B$. Among 16,416 ♂ progeny, 2 ♂♂ $r\ f^{3n}$ and 3 ♂♂ $f^{1}\ B$ were obtained, demonstrating that the f^{ex} represents f^{3n} and f^{1} coexisting on the same X chromosomes. Further, these results demonstrate that f^{3n} and f^{1} are pseudo-alleles.

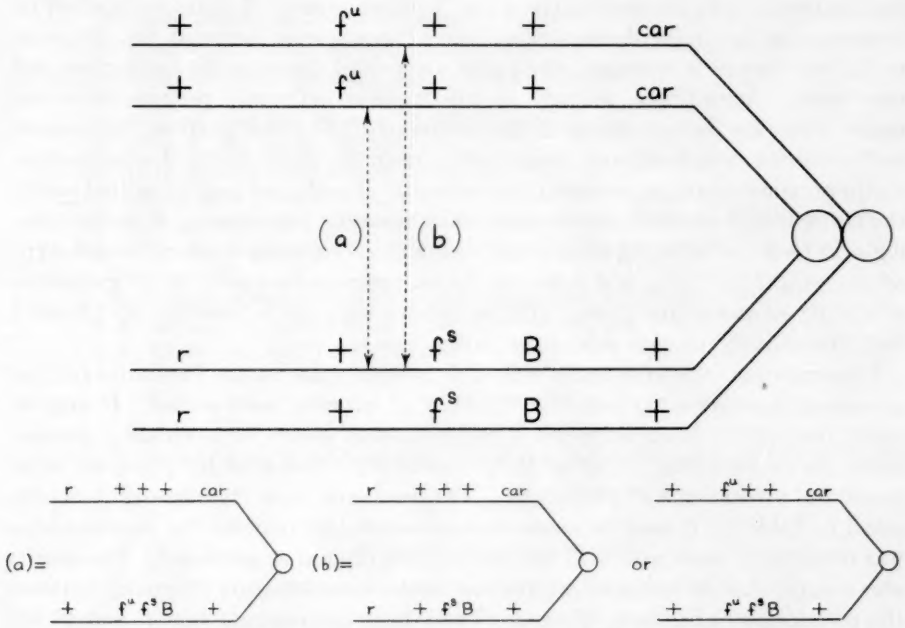


FIG. 1.—Scheme for detecting crossing-over between f pseudo-alleles in attached-X ♀♀. Upper diagram represents genotype of attached-X ♀♀ with f^{s} = suppressible and f^{u} = unsuppressible forked mutants. The reciprocal crossover is labeled (a), the nonreciprocal (b). The corresponding crossover products shown below are designated (a) and (b).

Since f^{5} is also suppressed by *su-f*, tests for recombination between f^{3n} and f^{5} were made. Attached-X ♀♀ of the genotype $r\ f^{5}\ B\ +/+\ f^{3n}\ +\ car$ were obtained and tested. The results are listed in Table 1. From these tests, one $f^{+}\ B/+$ ♀, derived from a reciprocal exchange between f^{3n} and f^{5}, was obtained. This conclusion was supported by the progeny test of the $f^{+}\ B/+$ ♀, the results of which are included in Table 2. Therefore, the $f^{+}\ B/+$ ♀ obtained was genotypically $r\ +\ +\ +/+\ f^{3n}\ f^{5}\ B$. Since f^{5} produces an extreme f phenotype, demonstration of the recovery of the $f^{3n}\ f^{5}\ B$ genotype followed the detachment of the X chromosomes of attached-X ♀♀ phenotypically $f^{ex}\ B$. On testing $f^{ex}\ B$ to *su-f*, it was found that

TABLE 2

PROGENY TESTS OF ♀♀ $^{+}B/+$

PHENOTYPE ♀ PROGENY	SOURCE OF ♀♀ Exp. 1	Exp. 2	Exp. 4
$f^{+}\ B/+$	78	207	40
$f^{ex}\ B$	6	7	1
$r\ f^{+}\ B^{+}$	7	4	2

f^{ex} is unsuppressed. Since f^5 is suppressed, the recovery of f^{ex} as unsuppressed follows if f^{3n} and f^5 coexist on the same X chromosome.

The crossing-over results listed in Table 1 indicate that recombination between the *f* pseudo-alleles is extremely rare and precludes the routine testing of recombination between independently occurring mutants by use of attached-X ♀ ♀. An attempt was made to obviate this difficulty in the following manner. The *f* locus is proximal to the centromere of the X chromosome. A centromere effect on crossing-over has been demonstrated,[5] i.e., crossing-over between loci proximal to the centromere is increased when they are moved distal to the centromere and vice versa. Accordingly, a study of crossing-over between *f* pseudo-alleles was made where the loci are distal to the centromere. For this purpose, the scute-8 (sc^8) inversion, which extends from a point near the distal tip to the heterochromatin proximal to the centromere, was selected. In sc^8 the *f* locus is shifted nearly the entire length of the X chromosome away from the centromere. For the crossing-over tests the following stocks were constituted: a tester stock of the genotype $sc^8\ B\ f^1/Cy,\ Ubx^{130}/Xa$; and stocks of the unsuppressed mutants of the genotypes $sc^8\ Bx^2\ f^{3n}\ un$ and $sc^8\ Bx^2\ f^3\ un$. (Bx^2 = Beadex wing, *un* = uneven eye, 2.4 and 1 unit, respectively, on each side of the *f* locus.)

Crossing-over tests were made with ♀ ♀ heterozygous for the *f* mutants and the autosomal inversions *Cy* and Ubx^{130}. Only ♂ progeny were scored. It may be noted that the products expected if recombination occurs between the *f* pseudo-alleles are the following: ♂♂ $sc^8\ Bx^2\ f^+$ and $sc^8\ B\ f^1\ f^{3n}\ un$ or $sc^8\ B\ f^1\ f^3\ un$, the latter genotypes producing a f^{ex} phenotype. The results of these crossing-over tests are listed in Table 3. It may be noted that recombination between the pseudo-alleles was obtained in each case with the markers distributed as predicted. The results also suggest that by using sc^8 an increase in the recombination frequency between the pseudo-alleles has been effected. These tests demonstrate that f^3 and f^{3n} are allelic.

TABLE 3

CROSSING-OVER BETWEEN *f* PSEUDO-ALLELES IN *In(1) sc*^{8} ♀ ♀

		PHENOTYPE EXCEPTIONAL ♂♂		TOTAL ♂♂
	GENOTYPE ♀ ♀	$sc^8\ B\ f^{ex}\ un$	$sc^8\ Bx^2\ f^+$	
1.	$sc^8\ B\ f^1/sc^8\ Bx^2\ f^{3n}\ un;\ Cy/+;\ Ubx^{130}/+$	3	0	16,440
2.	$sc^8\ B\ f^1/sc^8\ Bx^2\ f^3\ un;\ Cy/+;\ Ubx^{130}/+$	0	2	6,045

Discussion.—The foregoing analysis of four *f* mutants of *D. melanogaster* parallels in all essential details the earlier analyses made of the *v* eye-color mutants. Thus, by use of a suppressor, otherwise phenotypically similar *f* mutants have been separated into two classes, suppressible and unsuppressible. Morever, pseudo-allelism is indicated, since recombination occurred between the suppressed and unsuppressed mutants with all four mutants tested. Furthermore, there is evidence to indicate that cytologically the forked loci are associated with a doublet in the salivary gland chromosomes, a cytological feature which appears to be characteristic of pseudo-allelism.[6, 2] The forked locus has been assigned to region 1571-5 of the salivary gland X chromosome.[7] Within this region, a doublet consisting of bands 1571-2 occurs and, by analogy with the other instances of pseudo-allelism, represents, in all probability, the *f* loci.

The bearing that pseudo-allelism has on the gene concept has been discussed in detail elsewhere.[8] It will suffice here to point out that the *f* pseudo-alleles represent another instance where functional disparity is indicated, thereby rendering the existence of pseudo-alleles as parts of a larger "physiological" gene as unlikely.

Summary.—(1) By the use of a suppressor mutant, the forked bristle mutants of *D. melanogaster* can be classified into suppressible and unsuppressible mutants. (2) Pseudo-allelism is indicated, since recombination between suppressible and unsuppressible mutants has been demonstrated. (3) The significance of these results in relation to the gene concept is briefly considered.

[1] M. M. Green, these PROCEEDINGS, **38**, 300–305, 1952.
[2] M. M. Green, *ibid.*, **40**, 92–99, 1954.
[3] M. Whittinghell, *Genetics*, **23**, 300–306, 1938.
[4] E. B. Lewis, these PROCEEDINGS, **38**, 953–961, 1952.
[5] C. A. Offerman and H. J. Muller, *Proc. Sixth Intern. Cong. Genetics*, **2**, 143–145, 1932; G. W Beadle, these PROCEEDINGS, **18**, 160–165, 1932; A. H. Sturtevant and G. W. Beadle, *Genetics*, **21**, 554–604, 1936.
[6] E. B. Lewis, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 159–174, 1951.
[7] M. Demerec, *et al.*, *Carnegie Inst. Wash. Yearbook*, **41**, 191, 1942.
[8] M. M. Green, *Am. Nat.*, **89**, 65–71, 1955.

CURVILINEAR CLUSTER SETS OF ARBITRARY FUNCTIONS

BY F. BAGEMIHL

INSTITUTE FOR ADVANCED STUDY

Communicated by Marston Morse, April 4, 1955

We first prove a theorem concerning the structure of an arbitrary set of points in the unit circle $|z| < 1$ of the complex plane. Then we establish a property of the cluster sets of an arbitrary complex-valued function, defined in $|z| < 1$, along curves in $|z| < 1$ terminating in points of $|z| = 1$. Finally, we indicate several applications—in particular, to meromorphic functions in $|z| < 1$.

Let $K = \{z:|z| = 1\}$ and $D = \{z:|z| < 1\}$. We shall always understand θ to be restricted to the range $0 \leqq \theta < 2\pi$. By an *arc at* $e^{i\theta}$ we shall mean a Jordan arc in D having one end point at $e^{i\theta}$ and the other end point (regarded as belonging to the arc in question) at a point in D. The symbol P_θ will stand for the radius at $e^{i\theta}$: $\arg z = \theta$, $0 \leqq |z| < 1$. We shall denote the Riemann sphere by R and the closure of $M \subseteq R$ by $\bar{M}$. The function $f = f(z)$ will be supposed to be defined and single-valued in D and to assume values on R; as will be seen, we could take f to be more general, but we choose this as perhaps the most interesting case to which our results concerning cluster sets apply. We shall have occasion below to restrict f to be continuous, and we shall mean continuous in the extended sense, so that, e.g., a meromorphic function is to be regarded as continuous at a pole.

If Γ is an arc at $e^{i\theta}$, the cluster set of f at $e^{i\theta}$ along Γ, denoted by $C(f, \Gamma)$, is defined to be the set of all points $w \in R$ with the property that there exists a sequence of points $\{z_n\}$ on Γ for which $z_n \to e^{i\theta}$ and $f(z_n) \to w$ as $n \to \infty$. It follows readily that $C(f, \Gamma)$ is a nonempty closed set on R. If f is assumed to be continuous in D, then it is well known that $C(f, \Gamma)$ is a subcontinuum (possibly a single point) of R.

Modifying an idea due to Blumberg,[1] we prove

THEOREM 1. *If $S \subseteq D$, then there exists a $K^* \subseteq K$, with $K - K^*$ at most enumerable, such that, for every $e^{i\theta} \in K^*$, if Γ_1 and Γ_2 are arcs at $e^{i\theta}$, either Γ_1 and Γ_2 both intersect S or both intersect $D - S$.*

Proof: For every pair of rational numbers α, ρ satisfying

$$0 < \alpha < {}^1/_2\pi, \qquad 0 < \rho < 1, \tag{1}$$

we define a set $E_{\alpha,\rho} \subset K$; a point $e^{i\theta}$ shall belong to $E_{\alpha,\rho}$ if, and only if, there exist arcs Γ, Γ' at $e^{i\theta}$, with respective end points ζ, ζ' in D, such that

$$|\zeta| = |\zeta'| = \rho, \quad \theta - {}^1/_4\pi < \arg \zeta' < \theta + {}^1/_4\pi - \alpha < \arg \zeta < \theta + {}^1/_4\pi; \tag{2}$$

for every $z \in \Gamma \cup \Gamma'$ distinct from ζ and ζ',

$$\rho < |z| < 1, \qquad \theta - {}^1/_4\pi < \arg z < \theta + {}^1/_4\pi; \tag{3}$$

and

$$\Gamma \subseteq S, \qquad \Gamma' \subseteq D - S. \tag{4}$$

Define the set $E'_{\alpha,\rho}$ in the same way, except for the replacement of condition (4) by

$$\Gamma \subseteq D - S, \qquad \Gamma' \subseteq S. \tag{5}$$

Evidently $E_{\alpha,\rho}$ and $E'_{\alpha,\rho}$ are isolated sets and are therefore at most enumerable.

Let K^* be the set of all $e^{i\theta}$ that do not belong to any $E_{\alpha,\rho}$ or $E'_{\alpha,\rho}$, with α and ρ rational numbers satisfying relations (1). Since there are only enumerably many such sets $E_{\alpha,\rho}$, $E'_{\alpha,\rho}$, and each of these sets is at most enumerable, $K - K^*$ is at most enumerable.

Now suppose that $e^{i\theta} \in K^*$ and that Γ_1 and Γ_2 are arcs at $e^{i\theta}$. If Γ_1 and Γ_2 have at least one point in common, then the conclusion of Theorem 1 is obviously satisfied. Assume, then, that Γ_1 and Γ_2 do not intersect. Starting at $e^{i\theta}$, proceed along Γ_1 until a first point of $P_{\theta-{}^1/_4\pi} \cup P_{\theta+{}^1/_4\pi}$ is reached, and denote by Γ_1^* the arc at $e^{i\theta}$ thus described; in case Γ_1 intersects neither of the radii $P_{\theta\pm{}^1/_4\pi}$, put $\Gamma_1^* = \Gamma_1$. Define $\Gamma_2^* \subseteq \Gamma_2$ analogously. Choose a rational number ρ, between 0 and 1, so large that the circle $|z| = \rho$ intersects both Γ_1^* and Γ_2^* and contains in its interior the end points in D of these two arcs. Starting at $e^{i\theta}$, proceed along Γ_1^* until a first point, call it ζ, of $|z| = \rho$ is reached, and denote by Γ the arc at $e^{i\theta}$ thus described; proceeding along Γ_2^* from $e^{i\theta}$, define analogously the point ζ' and the arc Γ' at $e^{i\theta}$. Then $\zeta \neq \zeta'$, and there is clearly no loss of generality in assuming that, as viewed from the origin, ζ lies to the left of ζ' on $|z| = \rho$. On the minor arc of $|z| = \rho$ determined by the points $\rho e^{i(\theta\pm{}^1/_4\pi)}$, choose a point γ between ζ and ζ' such that $\theta + {}^1/_4\pi - \arg \gamma$ is a rational number—call this number α—between 0 and ${}^1/_2\pi$. Now we have relations (1) and (2), and relation (3) holds for every $z \in (\Gamma \cup \Gamma') - \{\zeta, \zeta'\}$. Neither condition (4) nor condition (5) is true, however; for otherwise we should have $e^{i\theta} \in E_{\alpha,\rho} \cup E'_{\alpha,\rho}$, a contradiction of our supposition that $e^{i\theta} \in K^*$. Hence, either Γ and Γ' both intersect S or both intersect $D - S$, which implies that the same is true of Γ_1 and Γ_2, Q.E.D.

With the aid of Theorem 1, we now prove

THEOREM 2. *If f is arbitrary, then there exists a $K^* \subseteq K$, with $K - K^*$ at most enumerable, such that, for every $e^{i\theta} \in K^*$ and every pair of arcs Γ_1, Γ_2 at $e^{i\theta}$,*

$$C(f, \Gamma_1) \cap C(f, \Gamma_2) \neq 0.$$

Proof: An open spherical cap of R, whose bounding circle is a rational distance from the center of the cap and whose center is either the point at infinity or a point whose stereographic projection in the z-plane has rational real and imaginary parts, will be called a rational cap. Enumerably many open sets on R possess the property of being expressible as the union of finitely many rational caps; let these open sets be $G_1, G_2, \ldots, G_n, \ldots$.

For every natural number n, let

$$S_n = \{z : z \in D, f(z) \notin G_n\}. \tag{6}$$

According to Theorem 1, there exists a $K_n^* \subseteq K$, with $K - K_n^*$ at most enumerable, such that, for every $e^{i\theta} \in K_n^*$, if Γ_1', Γ_2' are arcs at $e^{i\theta}$, then either Γ_1' and Γ_2' both intersect S_n or both intersect $D - S_n$. If we put

$$K^* = \bigcap_{n=1}^{\infty} K_n^*,$$

then $K - K^*$ is at most enumerable.

Now suppose that $e^{i\theta} \in K^*$ and that Γ_1 and Γ_2 are arcs at $e^{i\theta}$. Let us assume that

$$C(f, \Gamma_1) \cap C(f, \Gamma_2) = 0. \tag{7}$$

Then there exists an n such that

$$C(f, \Gamma_1) \subset G_n \tag{8}$$

and

$$C(f, \Gamma_2) \subset R - \bar{G}_n. \tag{9}$$

Let $\Gamma_1^1 \supset \Gamma_1^2 \supset \ldots \supset \Gamma_1^k \supset \ldots$ be a sequence of subarcs of Γ_1 at $e^{i\theta}$, whose end points in D converge to $e^{i\theta}$ as $k \to \infty$, and let $\Gamma_2^1 \supset \Gamma_2^2 \supset \ldots \supset \Gamma_2^k \supset \ldots$ be an analogous sequence of subarcs of Γ_2. Since $e^{i\theta} \in K_n^*$, it follows that, if k is a natural number, either Γ_1^k and Γ_2^k both intersect S_n or both intersect $D - S_n$. Hence there exists an increasing infinite sequence of natural numbers $\{k_j\}$ such that either, for every j, $\Gamma_1^{k_j}$ and $\Gamma_2^{k_j}$ both intersect S_n, or, for every j, both intersect $D - S_n$. In view of equation (6), the first alternative implies that $C(f, \Gamma_1) \cap (D - G_n) \neq 0$, which contradicts relation (8), and the second alternative implies that $C(f, \Gamma_2) \cap \bar{G}_n \neq 0$, which contradicts relation (9). Thus assumption (7) is untenable, and therefore Theorem 2 is true.

Remarks:

1. It is evident from the proofs of Theorems 1 and 2 that they are valid if the arcs Γ_1, Γ_2 at $e^{i\theta}$ are replaced by more general sets.

2. In Theorem 2 (and also in Theorem 1) the assertion that $K - K^*$ is at most enumerable cannot be strengthened, even if f is assumed to be regular in D. For suppose that $E = \{\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots\}$ is a given enumerable subset of K. Let $T_1, T_2, \ldots, T_n, \ldots$ be mutually external circles in $|z| \leq 1$ such that T_n is tangent to K at ζ_n, and let Γ_1^n, Γ_2^n be two chords of T_n terminating in ζ_n ($n = 1, 2, 3, \ldots$). Then there exists[2] a regular function $f(z)$ in D for which $\{0\} = C(f, \Gamma_1^n) \neq C(f, \Gamma_2^n) = \{1\}$ ($n = 1, 2, 3, \ldots$).

3. Theorems 1 and 2 can be transformed readily into theorems concerning a half-plane instead of D; and, if f is then taken to be arbitrary but real-valued instead of complex-valued, our theorems[3] sharpen or generalize results due to Blumberg,[4] Schmeiser,[5] and Jarník.[6]

Some applications of Theorem 2 are presented in the following corollaries.

COROLLARY 1. *Let f be arbitrary, and $E \subseteq K$. Suppose that to every $e^{i\theta} \in E$ there correspond two arcs, Γ_1^θ and Γ_2^θ, at $e^{i\theta}$, such that the two limits*

$$\lim_{\substack{z \to e^{i\theta} \\ z \in \Gamma_1^\theta}} f(z), \qquad \lim_{\substack{z \to e^{i\theta} \\ z \in \Gamma_2^\theta}} f(z)$$

exist and are unequal. Then E is at most enumerable.

COROLLARY 2. *If f is continuous in D, there exists a $K^* \subseteq K$, with $K - K^*$ at most enumerable, possessing the following property:*

For every $e^{i\theta} \in K^*$, let $\mathfrak{A}_\theta$ denote an arbitrarily given set of arcs at $e^{i\theta}$. Then,[7] for every $e^{i\theta} \in K^*$,

$$\bigcup_{\Gamma \in \mathfrak{A}_\theta} C(f, \Gamma)$$

is a connected set.

COROLLARY 3. *Let $A \subseteq K$ be an arc, $P \subseteq A$ be of positive measure on every subarc of A, $Q \subseteq A$ be of second category, $H \subseteq R$ be nonempty. Suppose that f is meromorphic and not identically constant in D; for every $e^{i\theta} \in P$ there is an arc Γ_θ at $e^{i\theta}$ such that $C(f, \Gamma_\theta) \subseteq H$; and, for every $e^{i\theta} \in Q$, $C(f, \Gamma_\theta) \neq R$. Then[8] H is of positive inner harmonic measure.*[9]

[1] H. Blumberg, "Exceptional Sets," *Fund. Math.*, **32**, 16–17, 1939.

[2] F. Bagemihl and W. Seidel, "Spiral and Other Asymptotic Paths, and Paths of Complete Indetermination, of Analytic and Meromorphic Functions," these PROCEEDINGS, **39**, 1251–1258, 1953, Theorem 1.

[3] To take account of the fact that in the real domain two values, $+\infty$ and $-\infty$, appear instead of the single point at infinity in the complex domain, a slight modification of the proof of Theorem 2 is required.

[4] Cf. H. Blumberg, *op. cit.*, pp. 17–18, Theorem IV and Corollary VII.

[5] *Ibid.*, p. 18, Corollary VIII.

[6] V. Jarník, "Sur les fonctions de deux variables réelles," *Fund. Math.*, **27**, 147, 1936. (Jarník remarks that his theorem also holds for complex-valued functions.)

[7] Cf. F. Hausdorff, *Grundzüge der Mengenlehre* (Leipzig, 1914), p. 245.

[8] Apply Theorem 7(b) (noting footnote 20) of F. Bagemihl and W. Seidel, "A General Principle Involving Baire Category, with Applications to Function Theory and Other Fields," these PROCEEDINGS, **39**, 1068–1075, 1953.

[9] I.e., H contains a closed subset of positive harmonic measure. (For the notion of harmonic measure see R. Nevanlinna, *Eindeutige analytische Funktionen* [2d ed.; Berlin-Göttingen-Heidelberg, 1953].)

*ON A CHARACTERIZATION OF THE NORMAL LAW**

By Glen Baxter

CORNELL UNIVERSITY

Communicated by S. Bochner, April 22, 1955

The following note contains a short proof of a theorem which Linnik[1] has recently given to characterize the normal law. In the proof we demonstrate the power in the concept of infinite divisibility in the generalized sense, a concept less widely known than we first had thought. Infinite divisibility in the generalized sense[2] has an advantage over ordinary infinite divisibility (although the two are equivalent) in that it may be much easier to verify in particular cases. This is true in the proof of the following:

THEOREM. *Let* $X_1 \not\equiv 0$ *and* X_2 *be independent, identically distributed random variables, and suppose that there exists a pair of real numbers* (a, b) *with* $0 < a \leqq b < 1$, $a^2 + b^2 = 1$, *for which* $aX_1 + bX_2$ *has the same distribution as* X_1. *Then* X_1 *is Gaussian with mean zero.*

Proof: If $\varphi(t)$ is the common characteristic function of X_1 and X_2, then

$$\varphi(at)\varphi(bt) = \varphi(t), \tag{1}$$

and, by iteration,

$$\varphi(a^n t)\varphi(a^{n-1}bt)^{\binom{n}{1}} \ldots \varphi(b^n t) = \varphi(t). \tag{2}$$

Now, for any fixed $\eta > 0$, there exists an integer n so large that $P(|a^{n-k}b^k X_1| \geqq \eta) \leqq \eta$ for $k = 0, 1, \ldots, n$. This shows that X_1 is infinitely divisible in the generalized sense and that it is possible to write

$$\log \varphi(t) \equiv \psi(t) = i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\, dG(x) \tag{3}$$

for some real constants γ, σ^2 and a bounded nondecreasing function $G(x)$ which is continuous when $x = 0$. Moreover, from equation (1)

$$\psi(at) + \psi(bt) = \psi(t). \tag{4}$$

Note that the real part of $\psi(t)$ can be written in the form

$$-\frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \frac{\cos ix - 1}{x^2}\, dF(x), \tag{5}$$

where

$$F(u) = \int_0^u (1 + x^2)\, dG(x). \tag{6}$$

$F(x)$ is monotone increasing, and $F(0) = \lim_{x \to 0} F(x) = 0$. Taking the real part of equation (4) and introducing expression (5), we find that

$$\int_{-\infty}^{\infty} \frac{\cos atx - 1}{x^2}\, dF(x) + \int_{-\infty}^{\infty} \frac{\cos btx - 1}{x^2}\, dF(x) = \int_{-\infty}^{\infty} \frac{\cos tx - 1}{x^2}\, dF(x), \tag{7}$$

which after a change of variables, becomes

$$\int_{-\infty}^{\infty} \frac{\cos tu - 1}{u^2} d\left[a^2F\left(\frac{u}{a}\right) + b^2F\left(\frac{u}{b}\right) - F(u)\right] = 0. \tag{8}$$

Hence

$$a^2F\left(\frac{u}{a}\right) + b^2F\left(\frac{u}{b}\right) = F(u), \tag{9}$$

and, by iteration,

$$\sum_{k=0}^{n} \binom{n}{k} a^{2n-2k}b^{2k}F\left(\frac{u}{a^{n-k}b^k}\right) = F(u). \tag{10}$$

For $u > 0$,

$$\begin{aligned} F(u) &\leqq F\left(\frac{u}{b^n}\right) = (a^2 + b^2)^nF\left(\frac{u}{b^n}\right) \\ &\leqq \sum_{k=0}^{n} \binom{n}{k} a^{2n-2k}b^{2k}F\left(\frac{u}{a^{n-k}b^k}\right) = F(u), \end{aligned} \tag{11}$$

and so equality must exist between the first two terms in equation (11). Since $F(u)$ is monotonic, it must be constant for all $u > 0$ and, moreover, for positive u, $F(u) = \lim_{x \to 0^+} F(x) = 0$. A similar argument shows that $F(u)$ is zero for negative u. From formulas (6) and (3),

$$\psi(t) = i\gamma t - \frac{\sigma^2t^2}{2}. \tag{12}$$

Finally, γ is necessarily zero if $\psi(t)$ is to satisfy equation (4), and the proof of the theorem is completed.

It was originally hoped that a generalization of the above theorem would yield a characterization of the stable laws. The obvious conjecture to make is that if X_1 and X_2 are independent, identically distributed (perhaps even symmetric) random variables, and if there exists a pair of real numbers (a, b) with $0 < a \leqq b < 1$, $a^\alpha + b^\alpha = 1$, for which $aX_1 + bX_2$ has the same distribution as X_1, then X_1 is stable with parameter α. It is clear that, if X_1 is stable with parameter α, i.e., if the characteristic function of X_1 is

$$\varphi(t) = \exp(-\lambda|t|^\alpha), \tag{13}$$

and if X_2 has the same distribution but is independent of X_1, then the hypotheses of the conjecture are satisfied for any pair of real numbers (a, b). Unfortunately, for $0 < \alpha < 2$ there exist examples of distributions which satisfy the hypotheses of the conjecture for particular values a and b but which are not stable.

Example: The distribution common to X_1 and X_2 will be defined in terms of its characteristic function. It is clear from the proof above that any distribution satisfying the hypotheses in the conjecture will be infinitely divisible in the generalized sense and so will have a Levy-Khintchine representation in the form of equation (3). For any fixed $0 < \alpha < 2$, let $\varphi_\alpha(t)$ be the characteristic function whose logarithm is given by equation (3), with $\gamma = \sigma^2 = 0$, and

$$G(x) = \begin{cases} \frac{1}{2} \sum_{k \leq \log_{\sqrt[\alpha]{2}} x} \frac{2^{[(2/\alpha)-1]k}}{1+2^{2k/\alpha}}, & x \geqq 0, \\ -G(-x), & x \leqq 0. \end{cases} \tag{14}$$

An easy calculation shows that

$$\log \varphi_\alpha(t) = \psi_\alpha(t) = \sum_{k=-\infty}^{\infty} \frac{\cos t2^{k/\alpha} - 1}{2^k}. \tag{15}$$

If $a = b = (1/2)^{1/\alpha}$, the hypotheses of the conjecture are certainly satisfied, but the distribution implied by equation (15) is not stable.

* This work was supported by the United States Air Force under contract No. AF18(600)-685, monitored by the Office of Scientific Research.

[1] Linnik, Yu. V., "Linear Statistics and the Normal Distribution Law," *Doklady Akad. Nauk SSSR* (N.S.), **83**, 353–355, 1952; "On Some Identically Distributed Statistics," *ibid.*, **89**, 9–11, 1953.

[2] Doob, J. L., *Stochastic Processes* (New York: John Wiley & Sons, 1953), chap. iii.

ON THE ADDITIVITY OF THE TRACE IN FINITE FACTORS

BY RICHARD V. KADISON*

COLUMBIA UNIVERSITY

Communicated by Paul A. Smith, March 17, 1955

1. *Introduction.*—The additivity of the trace has always seemed to those of us familiar with the technical details of the subject of rings of operators to be the most difficult single point in the basic theory. The proof of this fact given by Murray and von Neumann,[1] mysteriously enough, seems to require much more effort than so basic a point merits; and the fact that several other natural attempts to define the trace and establish its additivity lead to simple statements which appear to require the additive trace itself for their proofs only serves to deepen the mystery.

In the next section we present a completely natural and elementary proof of the additivity of the trace. In fact, our point of departure is the dimension theory of projections in a factor—we establish the existence of the trace and its additivity independently of the "semi-authenticated" trace.[2] Our proof was inspired by that of Murray and von Neumann,[1] the last paragraph of our proof being a direct adaptation, in modern dress, of their "local linear approximability" argument.[3]

2. *The Trace.*—If φ is a state of a factor M of type II_1 such that $\varphi(B^*B) = \varphi(BB^*)$ for each B in M, then φ is the trace on M. In fact, φ is the same on equivalent projections from the foregoing equality, and this, combined with the state properties of φ, implies that φ is the dimension function on projections in M; whence our assertion. It suffices, of course, to establish that $\varphi(B^*B) \leq \varphi(BB^*)$ for each B in M (substituting B^* for B, we conclude equality). If we can locate a state φ_n of M, for each positive integer n, such that $\varphi_n(B^*B) \leq (n + 1/n)\varphi_n(BB^*)$, for each B in M, then, by weak compactness of the set of states, we can choose a weak limit φ of some subsequence of (φ_n) which will of course satisfy $\varphi(B^*B) \leq \varphi(BB^*)$, for each B in M. Suppose now that we have located a positive, linear functional ρ

on M such that $\rho(A^*A) \leq (n + 1/n)\rho(AA^*)$, for each A in EME, with E some nonzero projection in M. We may assume that $D(E) = 1/m$, with m some positive integer (choosing a subprojection of E in place of E, if necessary). Let $E = E_1, \ldots, E_m$ be m, orthogonal, equivalent projections in M, and let V_i be a partial isometry in M such that $V_i^*V_i = E_1$, $V_iV_i^* = E_i$. We define φ_n by $\varphi_n(B) = \sum_{i=1}^{m} \rho(V_i^*BV_i)$. Then

$$\varphi_n(B^*B) = \sum_i \rho(V_i^*B^*(\sum_j E_j)\, BV_i) = \sum_{i,j} \rho(V_i^*B^*V_jV_j^*BV_i)$$

$$= \sum_{i,j} \rho[(E_1V_i^*B^*V_jE_1)(E_1V_i^*B^*V_jE_1)^*] \leq \frac{n+1}{n} \sum_{i,j} \rho[(V_j^*BV_i)\,(V_i^*B^*V_j)]$$

$$= \frac{n+1}{n} \sum_j \rho(V_j^*B(\sum_i E_i)B^*V_j) = \frac{n+1}{n}\,\varphi_n(BB^*).$$

Clearly φ_n is a positive, linear functional on M, and multiplying it by a positive scalar does not change its properties, so that we may assume that φ_n is a state.

It remains to locate a projection such as E and a positive, linear functional such as ρ. Suppose that we have found a positive, linear functional ρ and a nonzero projection E such that

$$D(F) \leq \rho(F) \leq \frac{n+1}{n}\, D(F) \qquad (*)$$

for each projection F in E. We write M in place of EME. Then we may replace F by an equivalent projection G in the left and right terms of (*) and conclude that

$$\rho(F) \leq \frac{n+1}{n}\, D(G) \leq \frac{n+1}{n}\, \rho(G). \qquad (**)$$

Of course, (**) holds if F and G are replaced by a positive, linear combination of mutually orthogonal projections and its transform via a unitary operator in M, respectively. The Spectral Theorem, uniform continuity of ρ, and this last remark imply that $\rho(A) \leq (n + 1/n)\rho(U^*AU)$, for each positive A and unitary U in M. In particular, $\rho(A^*A) \leq (n + 1/n)\rho(AA^*)$, since A^*A and AA^* are unitarily equivalent in M (use the polar decomposition of A, and recall that partial isometries in finite factors have unitary extensions).

We locate E and ρ such that (*) holds. Let η be any normal state of M (such as a vector state), and define $s(G)$ and $i(G)$ to be sup $\eta(F)/D(F)$ and inf $\eta(F)/D(F)$, respectively, as F ranges over the nonzero subprojections of G. There exists a nonzero subprojection F of G such that $i(F) \geq \eta(G)/D(G)$ (and, by precisely the same proof with each inequality sign reversed, there exists a subprojection F' such that $s(F') \leq \eta(G)/D(G)$). In fact, choose, by Zorn's Lemma, a maximal orthogonal family of projections (G_α) such that $\eta(G_\alpha)/D(G_\alpha) < \eta(G)/D(G)$. If (G_α) is void, G may be taken as F. If not, $G - \sum_\alpha G_\alpha \neq 0$, for otherwise

$$\frac{\eta(G)}{D(G)} = \frac{\Sigma\eta(G_\alpha)}{\Sigma D(G_\alpha)} < \frac{\Sigma D(G_\alpha)\eta(G)/D(G)}{\Sigma D(G_\alpha)} = \frac{\eta(G)}{D(G)},$$

a contradiction. Clearly $G - \sum_\alpha G_\alpha$ serves as the desired F in this case. Since

$\eta(I) = 1 = D(I)$, we can, from the above, find a nonzero projection G with $i(G) = a \geq 1$. Choose a nonzero projection F in G with $\eta(F)/D(F) \leq (n+1)a/n$. Once more, we can find a nonzero projection E in F such that $s(E) \leq (n+1)a/n$. Certainly $a \leqq i(E)$, so that, for each nonzero projection P in E, $D(P) \leq \eta(P)/a \leq (n+1)D(P)/n$. Taking ρ to be η/a, the proof is complete.

* The author is a Fulbright grantee.

[1] F. J. Murray and J. von Neumann, "On Rings of Operators. II," *Trans. Am. Math. Soc.*, **41**, 208–248, 1937; see, especially, pp. 210–217.

[2] F. J. Murray and J. von Neumann, "On Rings of Operators," *Ann. Math.*, **37**, 116–229, 1936.

[3] It has come to the author's attention that J. Dixmier has made an analogous remark in a forthcoming book on "rings of operators," which, incidentally, introduces many elegant simplifications in the subject.

REPRESENTATION OF A CLASS OF STOCHASTIC PROCESSES

BY SAMUEL KARLIN AND JAMES MCGREGOR

CALIFORNIA INSTITUTE OF TECHNOLOGY AND STANFORD UNIVERSITY

Communicated by S. Bochner, April 25, 1955

This note is a review of results obtained from a new approach to the study of certain time-homogeneous Markoff processes. The processes studied are of three types: (*a*) random walks, (*b*) birth and death processes, and (*c*) diffusion processes. Only the one-dimensional cases have been considered, although the method is not limited to these. The property which is common to the above processes and which makes them all amenable to the same treatment is the local nature of the changes that occur. This property can be characterized in terms of the semigroup of operators related to the process. Specifically, the infinitesimal generators of the semigroups in question are second-order differential or difference operators.

A random walk is the motion along a line executed by a particle which at each unit time can either stand still, or move one unit to the right, or move one unit to the left, the probabilities of these transitions depending perhaps on the position of the particle but not on the time. The set of possible positions (or states) of the particle is thus identifiable with a finite or infinite set of integers. The probability that if the particle starts at position i, it will be at position j after n units of time is denoted by $P_{ij}{}^{n}$. The fundamental matrix $P = (P_{ij}{}^{1})$ is given, and one wishes to know various statistical properties of the motion.

A birth and death process is essentially a random walk in which the time parameter has been made continuous. Let $P_{ij}(t)$ be the probability that if the particle was initially in state i, it will be in state j at time t. The process is called a birth and death process if, as $t \to 0$,

$$\begin{aligned} P_{i,\,i+1}(t) &= \lambda_i t + o(t), \\ P_{i,\,i-1}(t) &= \mu_i t + o(t), \\ P_{i,\,i}(t) &= 1 - (\lambda_i + \mu_i)t + o(t), \end{aligned}$$

where λ_i and μ_i are constants which may be thought of as the rates of absorption from state i into states $i+1$ and $i-1$.

Diffusion processes arise when both the state space and the time space are made continuous and the local change probabilities are described by Lindeberg-type conditions.

The basic idea of the approach presented here is to obtain integral representations for the transition probabilities in terms of a certain system of orthogonal functions. These orthogonal functions appear as eigenfunctions of an operator related to the process. The integral representations are then exploited to yield new probabilistic results and other consequences of analytical interest.

The simplest illustration is provided by the random walk on the nonnegative integers with fundamental matrix

$$P = \begin{bmatrix} r_0 & p_0 & & & 0 \\ q_1 & r_1 & p_1 & & \\ & q_2 & r_2 & p_2 & \\ & \cdot & \cdot & \cdot & \\ 0 & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & & \cdot \end{bmatrix},$$

where $p_i > 0$, $q_i > 0$, $r_i \geq 0$, and $p_i + r_i + q_i \leq 1$. In this case the orthogonal functions are the polynomials defined by the relations

$$\begin{aligned} Q_0(x) &\equiv 1, \\ xQ_0(x) &= r_0Q_0(x) + p_0Q_1(x), \\ xQ_n(x) &= q_nQ_{n-1}(x) + r_nQ_n(x) + p_nQ_{n+1}(x), \qquad n \geq 1. \end{aligned}$$

There is a unique positive regular measure ψ on $-1 \leq x \leq 1$, of total mass 1, with respect to which the polynomials form an orthogonal system. The integral representation of the transition probabilities is

$$P_{ij}^{\,n} = \pi_j \int_{-1}^{1} x^n Q_i(x) Q_j(x)\, d\psi(x),$$

where $\pi_0 = 1$ and $\pi_n = p_0p_1 \ldots p_{n-1}/q_1q_2 \ldots q_n$ for $n \geq 1$. This integral representation can be used to study recurrence and coincidence problems, absorption phenomena, and ergodic properties of the random walk.

A random walk on a finite set of integers leads to a similar representation in terms of a finite set of orthogonal polynomials. If the random walk takes place on all the positive and negative integers, the polynomials are replaced by 2×2 matrices with polynomial coefficients, although in special cases a representation in terms of polynomials orthogonal on the unit circle can be given.

To illustrate the scope and power of the method, the results for birth and death processes will be set forth in greater detail.

The transition probability matrix $P(t) = (P_{ij}(t))$, $t \geq 0$, of the birth and death process satisfies

$$\begin{aligned} &\text{I.} \quad P'(t) = P(t)A, \\ &\text{II.} \quad P'(t) = AP(t), \\ &\text{III.} \quad P(0) = \text{I}, \end{aligned}$$

where A is the infinite matrix

$$A = \begin{bmatrix} -(\lambda_0 + \mu_0) & \lambda_0 & 0 & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \end{bmatrix}.$$

It is assumed that $\lambda_i > 0$ for $i \geq 0$, $\mu_i > 0$ for $i \geq 1$, and $\mu_0 \geq 0$. The quantity μ_0 is interpreted as the rate of absorption from the zero state into a state of permanent death. Since there are always infinitely many matrices with properties I, II, and III, it is natural to consider additional restrictions associated with the probability model, namely,

$$\text{IV.} \quad P_{ij}(t) \geq 0,$$
$$\text{V.} \quad \sum_{j=0}^{\infty} P_{ij}(t) \leq 1,$$
$$\text{VI.} \quad P(t + s) = P(t)P(s).$$

Here again the orthogonal functions are polynomials and are defined by the relations

$$\begin{aligned} Q_0(x) &\equiv 1, \\ -xQ_0(x) &= -(\lambda_0 + \mu_0)Q_0(x) + \lambda_0 Q_1(x), \\ -xQ_n(x) &= \mu_n Q_{n-1}(x) - (\lambda_n + \mu_n)Q_n(x) + \lambda_n Q_{n+1}(x), \; n \geq 1. \end{aligned}$$

There is at least one positive regular measure ψ on $0 \leq x < \infty$, of total mass 1, with respect to which the polynomials form an orthogonal system. This is established by first noting that the polynomials define a sequence of moments and then showing that the corresponding Stieltjes moment problem has a solution. Any such *positive* measure ψ will be called a *solution of the moment problem.*

To each solution ψ of the moment problem a matrix $P(t)$ is defined by

$$\text{VII.} \quad P_{ij}(t) = \pi_j \int_0^{\infty} e^{-xt} Q_i(x)Q_j(x)\, d\psi(x),$$

where $\pi_0 = 1$ and $\pi_n = \lambda_0\lambda_1 \ldots \lambda_{n-1}/\mu_1\mu_2 \ldots \mu_n$ for $n \geq 1$. The matrix $P(t)$ has properties I, II, and III, and distinct solutions of the moment problem give rise to distinct matrices $P(t)$. This statement would remain true even if the requirement that ψ be positive were dropped, and it is seen that there are always infinitely many different matrices with analytic components and properties I, II, and III.

A solution of the moment problem is called *extremal* if the Parseval equation

$$\int_0^{\infty} |f(x)|^2\, d\psi(x) = \sum_{n=0}^{\infty} \left| \int_0^{\infty} f(x)Q_n(x)\, d\psi(x) \right|^2 \pi_n$$

is valid for every function f in $L_2(\psi)$. It is known that if the solution of the moment problem is unique, then it is extremal, and that if the solution of the moment problem is not unique, then there is a continuum of extremal solutions. The extremal solutions are important because of the following theorems.

THEOREM. *Let ψ be a solution of the moment problem and $P(t)$ the corresponding matrix defined by VII. Then $P(t)$ has properties IV and V. In order that $P(t)$ have property VI, it is necessary and sufficient that ψ be extremal.*

THEOREM. *Any matrix $P(t)$ with properties I–VI can be represented in the form VII, where ψ is an extremal solution of the moment problem.*

These theorems imply that formula VII establishes a one-to-one correspondence between the set of all matrices with properties I–VI and the set of all extremal solutions of the moment problem. The following uniqueness criterion replaces various previously known necessary but not sufficient and sufficient but not necessary conditions.

THEOREM. *In order that there be one and only one matrix $P(t)$ with properties I–V, it is necessary and sufficient that*

$$\sum_{n=0}^{\infty} \left(\pi_n + \frac{1}{\lambda_n \pi_n} \right)$$

be divergent.

It is known[1] that among the matrices with properties I–IV there is one, hereafter denoted by $M(t)$, which is minimal in the sense that if $P(t)$ is any other matrix with properties I–IV, then, for each i, j and all $t \geq 0$,

$$M_{ij}(t) \leq P_{ij}(t).$$

The minimal solution $M(t)$ also has properties V and VI and hence corresponds to an extremal solution ψ^* of the moment problem. The measure ψ^* can be obtained as the weak limit of a sequence of measures $\{\psi_n\}$, where ψ_n has only $n + 1$ points of increase, located at the zeros of $Q_{n+1}(x)$.

When the solution of the moment problem is not unique, formula VII provides a construction for infinitely many different matrices with properties I–V. If $P(t)$ is such a matrix distinct from $M(t)$, and if $c > 1$, then the matrix $M(t) + c\,[P(t) - M(t)]$ has properties I–IV but not V. The equality in V may fail to hold for $M(t)$ even if $M(t)$ is the only matrix with properties I–V. The present approach leads to a necessary and sufficient condition for the equality, which is a considerable sharpening of previous results in this direction.

THEOREM. *Let $M(t)$ be the minimal solution, and*

$$w_n = \frac{1}{\lambda_n} + \frac{\mu_{n+1}}{\lambda_n \lambda_{n+1}} + \ldots + \frac{\mu_{n+1} \ldots \mu_{n+k}}{\lambda_n \lambda_{n+1} \ldots \lambda_{n+k}} + \ldots .$$

In order that $\sum_{j=0}^{\infty} M_{ij}(t) = 1$ for each i and all $t \geq 0$, it is necessary and sufficient that $\sum_{n=0}^{\infty} w_n$ be divergent.

Certain ergodic theorems can be derived very easily from the integral representations. Let ψ be a solution of the moment problem and $P(t)$ the corresponding matrix. It can be seen by inspection of formula VII that $\lim_{t \to \infty} P_{ij}(t)$ exists. The limits are all positive if ψ has mass at $x = 0$, all zero otherwise. This result embodies a classical ergodic theorem. The following new and much deeper ergodic result is also easily obtained from formula VII.

THEOREM. *Let ψ be a solution of the moment problem and $P(t)$ be the corresponding matrix. Then, for each i, j, k, l,*

$$\lim_{t\to\infty} \frac{P_{ij}(t)}{P_{kl}(t)}$$

exists and is finite and not zero.

A fundamental property of the matrices which satisfy properties I–VI is their oscillating character, expressed in the following statement.

THEOREM. *Let $P(t)$ be any matrix with properties I–VI, and let $0 \leq i_1 < i_2 < \ldots < i_n, 0 \leq j_1 < j_2 < \ldots < j_n$. Then*

$$det \left| P_{i_\mu, j_\nu}(t) \right| > 0 \qquad for \qquad t > 0.$$

This theorem implies many interesting variation diminishing properties of $P(t)$ which will be elaborated elsewhere.

Birth and death processes with only a finite number of possible states lead to representations in terms of finite systems of orthogonal polynomials. Diffusion processes admit representations in terms of appropriate systems of orthogonal functions. It is worth noting that many of the well-known stochastic models lead to classical polynomials. For example, the telephone trunking model[2] leads to the Poisson-Charlier polynomials, and the continuous Ehrenfest model[3] uses Krawtchouk polynomials in its representation formula. In such cases the measure ψ is known at once. In many other cases of practical interest ψ may be computed by a procedure which Titchmarsh used in the theory of Sturm-Liouville problems.

Details of the above results and of other applications will appear in a separate publication.

[1] G. E. H. Reuter and W. Ledermann, *Proc. Cambridge Phil. Soc.*, **49**, 247–262, 1953.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications* (New York, 1950), p. 377.

[3] R. Bellman and T. Harris, *Pacific J. Math.*, **1**, 179–193, 1951.

---

*A VARIATIONAL METHOD IN THE THEORY OF HARMONIC INTEGRALS**

BY CHARLES B. MORREY, JR., AND JAMES EELLS, JR.

UNIVERSITY OF CALIFORNIA, BERKELEY, AND INSTITUTE FOR ADVANCED STUDY

Communicated by H. Whitney, March 18, 1955

1. In his first paper[1] on harmonic integrals on domains of Euclidean n-space E^n (with Cartesian co-ordinates), W. V. D. Hodge used a variational method to solve certain boundary-value problems. In order to carry over the theory to a compact orientable Riemannian manifold $\mathfrak{M}$, Hodge and subsequent writers have employed methods of integral equations (see Hodge,[2] Kodaira,[3] and de Rham and Kodaira[4]).

The present note summarizes the techniques and results of a variational method which can be applied to $\mathfrak{M}$; a complete exposition will appear elsewhere. In particular, we give the orthogonal decomposition theorem of Kodaira and include rather

complete information about the components of that decomposition (see sec. 6 below). Some of our results about the decomposition have been obtained independently and by different methods by K. O. Friedrichs.

2. A function f on an open set G in E^n is said to be of *class* C_μ^k $(0 < \mu \leq 1)$ if it possesses partial derivatives of order $\leq k$ which satisfy uniform Hölder conditions if $\mu < 1$ and uniform Lipschitz conditions if $\mu = 1$. We say that f is of *class* $\mathfrak{L}_2$ on G if f is square-integrable on G. The function f is of *class* $\mathfrak{P}_2$ on G if f is in $\mathfrak{L}_2$ on G and there are functions $g_1, \ldots, g_n$ in $\mathfrak{L}_2$ on G such that, if R is any cell interior to G, there is a sequence of Lipschitz functions $\{f_j\}$ on $\bar{R}$ such that

$$\lim_{j \to \infty} \int_R \left[(f_j - f)^2 + \sum_{k=1}^{n} \left(\frac{\partial f_j}{\partial x^k} - g_k \right)^2 \right] dx = 0. \tag{1}$$

Functions of class $\mathfrak{P}_2$ have been studied in detail elsewhere (see Calkin[5] and Morrey[6]) and have been used extensively in potential theory and the calculus of variations.

Riemannian n-manifolds $\mathfrak{M}$ of class C_μ^k are defined using co-ordinate transformations of class C_μ^k; the metric tensor g_{ij} for such a manifold will be of class C_μ^{k-1}. *We assume that* $\mathfrak{M}$ *is a compact Riemannian* n*-manifold (without boundary) of class* C_1^1 *at least.* Exterior differential forms of degree r (called r-forms, $r = 0, \ldots, n$) on $\mathfrak{M}$ are of *class* $\mathfrak{L}_2$, $\mathfrak{P}_2$, C^k, C^∞, *or* analytic if their components in every admissible co-ordinate system on $\mathfrak{M}$ have these respective properties.

Let $\mathfrak{L}_2^r$ denote the real Hilbert space obtained by completing the linear space of all Lipschitz r-forms on $\mathfrak{M}$ (r-forms on $\mathfrak{M}$ whose components satisfy a Lipschitz condition) with respect to the invariant inner product

$$(\omega, \eta) = \int_{\mathfrak{M}} \omega \wedge * \eta, \tag{2}$$

where $\omega \wedge * \eta$ denotes the exterior product of ω with the adjoint $(n - r)$-form $* \eta$ of η; for notation see de Rham and Kodaira.[4] Let $\mathfrak{P}_2^r$ denote the real Hilbert space obtained by completing the linear space of all Lipschitz r-forms on $\mathfrak{M}$ with respect to any of the inner products given in equation (3) below.

Let $\mathfrak{U}$ denote a covering of $\mathfrak{M}$ by co-ordinate patches $U_1, \ldots, U_Q$, whose domains in E^n are $G_1, \ldots, G_Q$. Then we define the inner product

$$((\omega, \eta))_{\mathfrak{U}} = (\omega, \eta) + \sum_{q=1}^{Q} \int_{G_q} \sum_{(i),\, \alpha} \omega_{(i)x^\alpha} \eta_{(i)x^\alpha} \, dx, \tag{3}$$

where $\omega_{(i)x^\alpha} = \partial \omega_{(i)} / \partial x^\alpha$; here and later $\alpha = 1, \ldots, n$ and $(i) = i_1 \ldots i_r$, with $i_1 < \ldots < i_r$. Any two such inner products are topologically equivalent.

We let d denote the exterior differential operator on $\mathfrak{M}$ and let δ denote its adjoint. It is easy to see that d is a bounded linear transformation from the whole of $\mathfrak{P}_2^{r-1}$ into $\mathfrak{L}_2^r$. Similarly, δ is a bounded linear transformation from $\mathfrak{P}_2^{r+1}$ into $\mathfrak{L}_2^r$. We let $\mathfrak{H}^r$ denote the subspace of $\mathfrak{P}_2^r$ of *harmonic fields* on $\mathfrak{M}$; i.e., $\omega \in \mathfrak{H}^r$ means that ω is an r-form in $\mathfrak{P}_2^r$ such that $d\omega = \delta\omega = 0$. If $\mathfrak{M}$ and ω were sufficiently differentiable, ω would be a harmonic field if and only if the Laplacian $\Delta\omega = d\delta\omega + \delta d\omega = 0$ (in general, see sec. 4 below).

3. The following convergence theorem is essentially due to Rellich:[7]

A. *Weak convergence in* $\mathfrak{P}_2^r$ *implies strong convergence in* $\mathfrak{L}_2^r$.

If we define the (weak) Dirichlet integral of an r-form by

$$D(\omega) = (d\omega, d\omega) + (\delta\omega, \delta\omega), \tag{4}$$

then the lower semicontinuity of the norm and the boundedness of d and δ imply

B. *The Dirichlet integral is lower semicontinuous with respect to weak convergence in* $\mathfrak{P}_2^r$.

If $D_0(\omega)$ denotes the Dirichlet integral in the case of a domain G in E^n, then an inspection of equation (4) shows that we may write

$$D_0(\omega) = \int_G \sum_{(i),\alpha} (\omega_{(i)x^\alpha})^2\, dx + 2 \sum_{(i),(j)} \sum_{\alpha,\beta} \int_G (\omega_{(i)x^\alpha}\, \omega_{(j)x^\beta} - \omega_{(i)x^\beta}\, \omega_{(j)x^\alpha}) dx. \tag{5}$$

The second integral vanishes for any ω of class $\mathfrak{P}_2$ with support interior to G. That fact allows us to give a simple proof of the following lemma, essentially due to Gaffney:[8]

With each point p of $\mathfrak{M}$, each $\epsilon > 0$ and $r = 0, \ldots, n$, we may associate a constant L and a co-ordinate patch U containing p with domain $B_R(|x| < R)$ such that

$$D(\omega) \geq (1 - \epsilon) \int_{B_R} \sum_{(i),\alpha} (\omega_{(i)x^\alpha})^2\, dx - L(\omega, \omega) \tag{6}$$

for every $\omega \in \mathfrak{P}_2^r$ with support in U.

By using a partition of unity, we may easily show

C. *With each finite covering $\mathfrak{U}$ of $\mathfrak{M}$ by co-ordinate patches and each $r = 0, \ldots, n$, we can associate a pair of constants $K_\mathfrak{U}$ and $L_\mathfrak{U}$ such that*

$$D(\omega) \geq K_\mathfrak{U}((\omega, \omega))_\mathfrak{U} - L_\mathfrak{U}(\omega, \omega) \tag{7}$$

for all $\omega \in \mathfrak{P}_2^r$.

4. Using A, B, and C, we may now apply our variational method.

LEMMA 1. *Suppose that M is any closed linear manifold in $\mathfrak{L}_2^r$ such that $M \cap \mathfrak{P}_2^r$ is not empty. Then there is a form ω in $M \cap \mathfrak{P}_2^r$ which minimizes $D(\omega)$ among all such forms with $(\omega, \omega) = 1$.*

For let $\{\omega_k\}$ be a minimizing sequence, and take a co-ordinate covering $\mathfrak{U}$ of $\mathfrak{M}$. From relation (7) we see that $((\omega_k, \omega_k))_\mathfrak{U}$ is uniformly bounded. Hence a subsequence converges weakly in $\mathfrak{P}_2^r$ to some form ω in $\mathfrak{P}_2^r$. The lemma now follows from A and B.

THEOREM 1. *For each $r = 0, \ldots, n$ the linear manifold $\mathfrak{H}^r$ of harmonic fields on $\mathfrak{M}$ is finite dimensional.*

We choose linear manifolds $M_1, M_2, \ldots$ in $\mathfrak{L}_2^r$ and the minimizing forms ω_k in M_k $(k = 1, 2, \ldots)$, where $M_1 = \mathfrak{L}_2^r$, M_2 is orthogonal to ω_1, M_3 is orthogonal to ω_1 and ω_2, etc. Then $0 \leq D(\omega_1) \leq D(\omega_2) \leq \ldots$. Clearly ω_k is a harmonic field if and only if $D(\omega_k) = 0$. Suppose that this held for all k; then relation (7) would imply that $((\omega_k, \omega_k))_\mathfrak{U}$ is uniformly bounded, whence a subsequence $\{\omega_j\}$ would converge weakly in $\mathfrak{P}_2^r$ and hence strongly in $\mathfrak{L}_2^r$ to some ω_0. This is impossible, since the ω_j form an orthonormal set in $\mathfrak{L}_2^r$.

THEOREM 2. *For any co-ordinate covering $\mathfrak{U}$ of $\mathfrak{M}$ there is a constant $\lambda_\mathfrak{U} > 0$ such that, if ω is in $\mathfrak{P}_2^r$ and ω is orthogonal to $\mathfrak{H}^r$, then*

$$D(\omega) \geq \lambda_\mathfrak{U}((\omega, \omega))_\mathfrak{U}.$$

From Theorem 1 it follows that $\mathfrak{H}^r$ is spanned by a finite number $\omega_1, \ldots, \omega_k$ of the above sequence. For any ω in $\mathfrak{P}_2{}^r$ and orthogonal to $\mathfrak{H}^r$, it is easy to see from the homogeneity of $D(\omega)$ that

$$D(\omega) \geq D(\omega_{k+1}) \quad (\omega, \omega) = h(\omega, \omega) \qquad (h > 0). \quad (8)$$

The result now follows from relation (7).

THEOREM 3. *If ω_0 is any form in $\mathfrak{L}_2{}^r$ which is orthogonal to $\mathfrak{H}^r$, then there is a unique form Ω_0 in $\mathfrak{P}_2{}^r$ which is orthogonal to $\mathfrak{H}^r$ which satisfies*

$$(d\Omega_0, d\zeta) + (\delta\Omega_0, \delta\zeta) = (\omega_0, \zeta) \qquad (9)$$

for all $\zeta \in \mathfrak{P}_2{}^r$. Ω_0 is called the potential of ω_0.

By setting $\zeta = \Omega_1 - \Omega_0$, it is easy to see that Ω_0 is unique. Ω_0 is found by minimizing

$$I(\omega) = D(\omega) - 2(\omega, \omega_0) \qquad (10)$$

among all ω in $\mathfrak{P}_2{}^r$ which are orthogonal to $\mathfrak{H}^r$. To do this, it is sufficient to select minimizing sequences $\{\omega_k\}$ for which $I(\omega_k) \leq 0$; clearly, by Theorem 2, $((\omega_k, \omega_k))_{\mathfrak{U}}$ is uniformly bounded for any such sequence.

All the results above hold on any manifold of class $C_1{}^1$ or better. In case $\mathfrak{M}$, Ω_0, ω_0 are sufficiently differentiable, equation (9) implies that $\Delta\Omega_0 = \omega_0$. However, one sees easily that the *second* derivatives of the g_{ij}, as well as those of the components of Ω_0, enter into the formulas for $\Delta\Omega_0$ if $r > 0$; in our most general case none of these derivatives is at hand.

5. The following results concerning the differentiability of the potentials follow from the recent work of one of the writers on the solutions of elliptic systems of differential equations (see Morrey[9]). In the case at hand the system is *strongly elliptic*, and its special nature permits a great simplification of the general theory.

We say that ω is of *class* $\mathfrak{L}_{2\lambda}$ $(0 < \lambda < n/2)$ if, for each co-ordinate patch U with domain B_R, there is a constant $L_{U,\omega}$ such that

$$\int_{B_r} \sum_{(i)} |\omega_{(i)}|^2 \, dx \leq L^2{}_{U,\omega} \, r^{2\lambda} \qquad (0 \leq r \leq R); \quad (11)$$

$\mathfrak{P}_{2\lambda}$ is defined similarly. These spaces are important in the study of the growth of the Dirichlet integral.

THEOREM 4. *Suppose that ω is in $\mathfrak{L}_2{}^r$ and orthogonal to $\mathfrak{H}^r$, and let Ω be its potential.*

i) If $\mathfrak{M}$ is of class $C_1{}^1$, then Ω, $d\Omega$, and $\delta\Omega$ are in $\mathfrak{P}_2$.

ii) If $\mathfrak{M}$ is of class $C_1{}^1$ and ω is in $\mathfrak{L}_{2\lambda}$, then Ω, $d\Omega$, and $\delta\Omega$ are in $\mathfrak{P}_{2\lambda}$, and in $C_\mu{}^0$ if $\lambda = n/2 - 1 + \mu$ $(0 < \mu < 1)$.

iii) If $\mathfrak{M}$ is of class $C_\mu{}^k$ and ω is in $C_\mu{}^{k-2}$ $(k \geq 2, 0 < \mu < 1)$, then Ω, $d\Omega$, and $\delta\Omega$ are in $C_\mu{}^{k-1}$. If $k \geq 3$ and ω is in $C_\mu{}^{k-3}$, then Ω is in $C_\mu{}^{k-1}$.

iv) If $\mathfrak{M}$ and ω are of class C^∞ or are analytic, then so is Ω.

In all cases, if we set $\alpha = d\Omega$ and $\beta = \delta\Omega$, we have

$$\delta\alpha = d\beta = \delta(d\Omega) + d(\delta\Omega) = \omega, \qquad (12)$$

and $d\alpha = \delta\beta = 0$.

It should be observed that we can take the particular combination of second derivatives indicated in equation (12), even though Ω may not possess all second derivatives.

THEOREM 5. *Suppose that H is a harmonic field.*

i) If $\mathfrak{M}$ is of class C_1^1, then H, dH, and δH are in $\mathfrak{P}_{2\lambda}$ and in C_μ^0 ($\lambda = n/2 - 1 + \mu$) for any μ ($0 < \mu < 1$).

ii) If $\mathfrak{M}$ is of class C_μ^k ($0 < \mu < 1$, $k \geq 2$), then H, dH, and δH are in C_μ^{k-1}.

iii) If $\mathfrak{M}$ is of class C^∞ or is analytic, then so is H.

In both Theorem 4 and Theorem 5 the 0-forms have one additional degree of differentiability.

6. The following result complements the well-known orthogonal decomposition theorem of Kodaira[3] (for compact Riemannian manifolds of class at least C_1^1).

THEOREM 6. *If ω is any form in $\mathfrak{L}_2$, there is a harmonic field H and forms α, β, and Ω in $\mathfrak{P}_2$ such that*

$$\begin{aligned} \omega &= H + \delta\alpha + d\beta, \qquad d\alpha = \delta\beta = 0, \\ \alpha &= \qquad\quad d\Omega, \qquad \beta = \delta\Omega, \end{aligned} \tag{13}$$

where Ω is the potential of $\omega - H$. If the first of equations (13) *also holds for a harmonic field H_1 and for forms α_1 and β_1 in $\mathfrak{P}_2$, then $\delta\alpha_1 = \delta\alpha$, $d\beta_1 = d\beta$, and $H_1 = H$.*

If ω is in $\mathfrak{P}_2$, then α and β are the potentials of $d\omega$ and $\delta\omega$, respectively, whence $\delta\alpha$ and $d\beta$ are in $\mathfrak{P}_2$.

The sets $\mathfrak{C}^r$ of all forms $\delta\alpha$ for α in $\mathfrak{P}_2{}^{r+1}$ and $\mathfrak{D}^r$ of all forms $d\beta$ for β in $\mathfrak{P}_2{}^{r-1}$ are closed linear manifolds in $\mathfrak{L}_2{}^r$, and we have

$$\mathfrak{L}_2{}^r = \mathfrak{H}^r \oplus \mathfrak{C}^r \oplus \mathfrak{D}^r.$$

The differentiability properties of H have been stated in Theorem 5; those of $\delta\alpha$ and $d\beta$ may be determined from Theorems 4 and 6 when ω and $\mathfrak{M}$ have further differentiability properties.

* This research was partially supported by the United States Air Force, through the Office of Scientific Research of the Air Force Research and Development Command. Presented to the American Mathematical Society, February 26, 1955.

[1] W. V. D. Hodge, *Proc. London Math. Soc.*, (ser. 2), **36**, 257–303, 1934.

[2] W. V. D. Hodge, *The Theory and Applications of Harmonic Integrals* (2d ed.; Cambridge: At the University Press, 1952).

[3] K. Kodaira, *Ann. Math.*, **50**, 587–665, 1949.

[4] G. de Rham and K. Kodaira, *Harmonic Integrals* (mimeographed notes, Institute for Advanced Study, 1950).

[5] J. W. Calkin, *Duke Math. J.*, **6**, 170–185, 1940.

[6] C. B. Morrey, Jr., *Duke Math. J.*, **6**, 186–215, 1940.

[7] F. Rellich, *Nachr. Ges. Wiss. Göttinger, Math.-physik. Kl.*, pp. 30–35, 1930.

[8] M. P. Gaffney, these PROCEEDINGS, **37**, 48–50, 1951.

[9] C. B. Morrey, Jr., *Ann. Math. Studies*, No. 33 (Princeton, N. J.: Princeton University Press, 1954).

ON THE CHARACTERS OF THE SYMMETRIC GROUP

By Francis D. Murnaghan

INSTITUTO TECNOLÓGICO DE AERONÁUTICA, SÃO JOSÉ DOS CAMPOS, BRASIL*

Communicated April 6, 1955

We have given in a previous note[1] formulas furnishing the characters of the irreducible representations of the symmetric group S_m on m symbols whose characteristics are $\{m-p, (\mu)\}$ for all partitions (μ) of any nonnegative integer $p \leq 5$, and we have later[2] furnished the formulas for $p = 6$ and $p = 7$. We have only now noticed a certain feature of the structure of these formulas which makes their determination almost trivial and their extension to higher values of p simple. When $p = 1$, the only partition of p is the partition (1), and the corresponding formula is $\alpha_1 - 1$, α_1 being the number of unary cycles in the class $(\alpha) = (\alpha_1, \ldots, \alpha_m)$ for which the character of the representation of S_m whose characteristic is $\{m-1, 1\}$ is furnished by the formula $\alpha_1 - 1$. We observe that $\alpha_1 - 1$ is the result of replacing m by α_1 in the expression $m - 1$ for the dimension of the irreducible representation of S_m whose characteristic is $\{m-1, 1\}$, and so we denote $\alpha_1 - 1$ by $d_{(1)}$. In general, we denote by $d_{(\mu)}$ the result of replacing m by α_1 in the expression for the dimension of the irreducible representation of S_m whose characteristic is $\{m-p, (\mu)\}$, so that, for example:

$$d_{(2)} = \frac{1}{2}\alpha_1(\alpha_1 - 3); \qquad d_{(1^2)} = \frac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2);$$

$$d_{(3)} = \frac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5); \qquad d_{(21)} = \frac{1}{3}\alpha_1(\alpha_1 - 2)(\alpha_1 - 4); \qquad d_{(1^3)} = \frac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3);$$

$$d_{(4)} = \frac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 7); \qquad d_{(31)} = \frac{1}{8}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1 - 6);$$

$$d_{(2^2)} = \frac{1}{12}\alpha_1(\alpha_1 - 1)(\alpha_1 - 4)(\alpha_1 - 5); \qquad d_{(21^2)} = \frac{1}{8}\alpha_1(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 5);$$

$$d_{(1^4)} = \frac{1}{24}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4).$$

We denote the fact that the character of the irreducible representation of S_m, whose characteristic is $\{m-1, 1\}$, is $\alpha_1 - 1 = d_{(1)}$ by the symbol $\{1\} : d_{(1)}$. This single formula suffices to construct the character table for S_3, since $\{0\} : 1$; more than this, it furnishes the first 2 rows, and the last 2 rows of the character table of S_m for every value of m. The formula for $\{2\}$ is a polynomial function of α_1 and α_2 of which the part independent of α_2 is $d_{(2)}$, while the part which involves α_2 is $d_{(0)}\alpha_2 = \alpha_2$. The coefficient of α_2 is obtained as follows: Subtract 2 from the single part 2 of $\{2\}$, obtaining $\{0\}$; then the coefficient of α_2 is $d_{(0)} = 1$. To obtain the coeffi-

cient of α_2 in the formula for $\{1^2\}$, subtract 2 from each of the two parts of 1^2, obtaining $\{-1,1\} + \{1, -1\} = -\{0\}$. Then the coefficient of α_2 in the desired formula is $-d_{(0)} = -1$. Thus

$$\{2\} : d_{(2)} + \alpha_2; \qquad \{1^2\} : d_{(1^2)} - \alpha_2.$$

These suffice, in conjunction with $\{0\} : 1$, $\{1\} : d_{(1)}$, to construct the character tables of S_4 and S_5; more than this, they furnish the first 4 rows and the last 4 rows of the character table of S_m for every value of m. We observe that the coefficients of α_2, 1 and -1, in the formulas for $\{2\}$ and $\{1^2\}$ are the numbers in the second column of the character table for S_2.

The formula for each of the partitions (μ) of 3 is a polynomial function of α_1, α_2, α_3 in which the term independent of α_2 and α_3 is, in each case, $d_{(\mu)}$. The coefficients of α_3 are the numbers 1, -1, 1 in the third column of the character table for S_3, and the coefficients of α_2 in the formulas for $\{3\}$, $\{21\}$, and $\{1^3\}$ are $d_{(1)}$, 0, and $-d_{(1)}$, respectively, since $\{3\} \to \{1\}$, $\{21\} \to \{01\} + \{2, -1\} = 0$, $\{1^3\} \to \{-1,1^2\} + \{1, -1,1\} + \{1,1, -1\} = -\{1\}$. Thus

$$\{3\} : d_{(3)} + d_{(1)}\alpha_2 + \alpha_3; \qquad \{21\} : d_{(21)} - \alpha_3; \qquad \{1^3\} : d_{(1^3)} - d_{(1)}\alpha_2 + \alpha_3.$$

These suffice, in conjunction with the formulas already given, to construct the character table for S_6 and to determine the first 7 rows and the last 7 rows of the character table of S_m for every value of m. We observe that the coefficients of $d_{(1)}\alpha_2$ are the numbers 1, 0, -1 of the second column of the character table of S_3.

For the partitions (μ) of 4 we encounter terms which correspond to the class of S_4 which has 2 binary cycles, and for these we must subtract 2 twice to obtain the coefficient of $\binom{\alpha_2}{2} = \frac{1}{2}\alpha_2(\alpha_2 - 1)$. For example, in deriving the formula for $\{21^2\}$, the first subtraction yields $\{01^2\} + \{2, -1,1\} + \{2,1, -1\} = -\{2\}$, and the second subtraction yields $-\{0\}$, so that the coefficient of $\binom{\alpha_2}{2}$ is -1. Thus

$$\{4\} : d_{(4)} + d_{(2)}\alpha_2 + d_{(1)}\alpha_3 + \binom{\alpha_2}{2} + \alpha_4;$$

$$\{31\} : d_{(31)} + d_{(1^2)}\alpha_2 - \binom{\alpha_2}{2} - \alpha_4;$$

$$\{2^2\} : d_{(2^2)} + (d_{(2)} - d_{(1^2)})\alpha_2 - d_{(1)}\alpha_3 + 2\binom{\alpha_2}{2};$$

$$\{21^2\} : d_{(21^2)} - d_{(2)}\alpha_2 - \binom{\alpha_2}{2} + \alpha_4;$$

$$\{1^4\} : d_{(1^4)} - d_{(1^2)}\alpha_2 + d_{(1)}\alpha_3 + \binom{\alpha_2}{2} - \alpha_4.$$

These suffice, in conjunction with the formulas already given, to construct the character table for S_7 and to determine the first 12 rows and the last 12 rows of the character table of S_m for every value of m. We observe that the coefficients of α_4 are the numbers of the last column of the character table of S_4, that the coefficients of $\binom{\alpha_2}{2}$ are the numbers of the fourth column of the character table of S_4, and that

the coefficients of $d_{(1)}\alpha_3$ are the numbers of the third column of the character table of S_4. We conclude by writing the formula for $\{6\}$:

$$\{6\}: d_{(6)} + d_{(4)}\alpha_2 + d_{(3)}\alpha_3 + d_{(2)}\alpha_4 + d_{(1)}\alpha_5 + d_{(2)}\binom{\alpha_2}{2} + d_1\alpha_2\alpha_3 + \binom{\alpha_2}{3} + \alpha_2\alpha_4 + \binom{\alpha_3}{2} + \alpha_6.$$

* Present address: 6202 Sycamore Road, Baltimore 12, Md.

[1] F. D. Murnaghan, these PROCEEDINGS, **37**, 55–58, 1951.

[2] F. D. Murnaghan, *Anais Acad. brasil. ciencias*, **23**, 141–154, 1951.

---

*CLUSTER SETS OF FUNCTIONS MEROMORPHIC IN THE UNIT CIRCLE**

BY KIYOSHI NOSHIRO

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY, AND NAGOYA UNIVERSITY, JAPAN

Communicated by Lars V. Ahlfors, March 25, 1955

The purpose of this note is to state some theorems which generalize recent results of Ohtsuka,[1] Lohwater,[2] Lehto,[3] and Storvick.[4]

1. Let $w = f(z)$ be a meromorphic function in the unit circle D: $|z| < 1$. Let $z_0 = e^{i\theta_0}$ be a fixed point on C: $|z| = 1$ and A an open arc of C containing z_0. We suppose that E is a set of measure zero containing z_0 and contained in A. We associate with every $e^{i\theta} \in A - E = \mathfrak{C}E$ an arbitrary simple curve $\Lambda(e^{i\theta})$ in D terminating at $z = e^{i\theta}$ and *the cluster set* $S_\Lambda(e^{i\theta})$ *of* $f(z)$ *at* $z = e^{i\theta}$ *with respect to* $\Lambda(e^{i\theta})$, which is defined in the following way: $\alpha \in S_\Lambda(e^{i\theta})$ if there is a sequence of points z_n $(n = 1, 2, \ldots)$ in D such that $f(z_n) \to \alpha$, $z_n \to e^{i\theta}$, $z_n \in \Lambda(e^{i\theta})$. Clearly $S_\Lambda(e^{i\theta})$ is either a continuum or a single point. We define the (interior) cluster set $S_{z_0}^{(D)}$, the (boundary) cluster set $S_{z_0}^{(C)}$, and the range of values $R_{z_0}^{(D)}$ of $f(z)$ as usual.[5] Further, we introduce another boundary cluster set $\Gamma_{z_0}^{(C)}$ of $f(z)$ at z_0 as follows: $\Gamma_{z_0}^{(C)} = \bigcap_{r>0} M_r$, where M_r denotes the closure of the union $\bigcup S_\Lambda(e^{i\theta})$ for all $e^{i\theta}$ in the intersection of $\mathfrak{C}E$ and $|z - z_0| < r$. It is clear that $S_{z_0}^{(D)}$, $S_{z_0}^{(C)}$, and $\Gamma_{z_0}^{(C)}$ are all nonempty closed sets and that $\Gamma_{z_0}^{(C)} \subset S_{z_0}^{(C)} \subset S_{z_0}^{(D)}$. It should be noted that in many cases $S_{z_0}^{(D)} - S_{z_0}^{(C)}$ becomes empty, while $S_{z_0}^{(D)} - \Gamma_{z_0}^{(C)}$ is not empty. Accordingly, it seems necessary to investigate interrelations among $S_{z_0}^{(D)}$, $\Gamma_{z_0}^{(C)}$, and $R_{z_0}^{(D)}$, applying classical results on cluster sets.[6]

2. We first state

THEOREM 1. $S_{z_0}^{(D)} - \Gamma_{z_0}^{(C)}$ *is an open set.*

Proof: Let w_0 be a point belonging to $S_{z_0}^{(D)} - \Gamma_{z_0}^{(C)}$. We first select a positive number r such that w_0 is outside M_r. Denote by ρ' the distance of w_0 from M_r. Next we select two points $z_1 = e^{i\theta_1}$, $z_2 = e^{i\theta_2}$ $(\theta_1 < \theta_0 < \theta_2)$ belonging to $\mathfrak{C}E \cap (K)$, where (K) denotes the disk $|z - z_0| < r$. We take a suitable last part $\Lambda'(e^{i\theta_1})$ of $\Lambda(e^{i\theta_1})$ and a suitable last part $\Lambda'(e^{i\theta_2})$ of $\Lambda(e^{i\theta_2})$ lying in $D \cap (K)$. Then it is possible to construct a cross-cut L of $D \cap (K)$ by connecting two end points, lying in D, of $\Lambda'(e^{i\theta_1})$ and $\Lambda'(e^{i\theta_2})$ by a rectilinear segment. We may suppose that $f(z) \neq w_0$ on L and that there exists a positive number ρ'' such that $|f(z) - w_0| \geqq \rho'' > 0$ on

L. Let ρ be a positive number less than min (ρ', ρ''). Denote by D_0 the region inclosed by the simple closed curve consisting of L and the circular arc $(e^{i\theta}; \theta_1 \leqq \theta \leqq \theta_2)$. Since w_0 is a cluster value of $f(z)$ at $z_0 = e^{i\theta_0}$, there is a sequence of points $z_\mu \in D_0$, $z_\mu \to z_0$, such that $f(z_\mu) \to w_0$. We keep this sequence z_μ $(\mu = 1, 2, \ldots)$ fixed throughout the proof. The inverse image of (c): $|w - w_0| < \rho$ in D_0 consists of at most a countable number of connected regions. We denote the component containing z_μ by Δ_μ which may conceivably coincide with other Δ_ν's.

We first treat the case in which there is an infinite number of components Δ_μ. In this case we assume for simplicity that $\Delta_\mu \neq \Delta_\nu$ for $\mu \neq \nu$. It is easily shown that Δ_μ $(\mu = 1, 2, \ldots)$ converges to z_0. If Δ_μ is compact in D_0 and if we denote by $f(\Delta_\mu)$ the set of values taken by $f(z)$ in Δ_μ, we have $(c) = f(\Delta_\mu)$. Suppose that Δ_μ is noncompact in D_0. It should be noted that Δ_μ is not necessarily simply connected. The boundary of Δ_μ consists of at most a countable number of analytic curves γ_μ (the relative boundary of Δ_μ) and a closed set on $|z| = 1$. Obviously $|f(z) - w_0| = \rho$ on γ_μ, and $|f(z) - w_0| < \rho$ in Δ_μ. We map the universal covering surface $\tilde{\Delta}_\mu$ of Δ_μ conformally upon the unit disk $|\zeta| < 1$ in a one-to-one manner. Denote by $z = z(\zeta)$ the mapping function. Considering the two bounded analytic functions $z = z(\zeta)$ and $w = w(\zeta) = f[z(\zeta)]$ in $|\zeta| < 1$, we denote by E_ζ the set of $\zeta = e^{i\varphi}$ such that both radial limits $z(e^{i\varphi})$, $w(e^{i\varphi})$ exist and $|w(e^{i\varphi}) - w_0| < \rho$. E_ζ is measurable. We consider first the case that there exists a point $e^{i\varphi_\mu} \in E_\zeta$ such that $z(e^{i\varphi_\mu})$ coincides with some $z = e^{i\theta_\mu} \in \mathfrak{C}E$. There exists an asymptotic path λ_μ terminating at $e^{i\theta_\mu}$ along which $f(z)$ converges to the asymptotic value $w(e^{i\varphi_\mu})$ lying in (c). Then, by a classical theorem on cluster sets, $\overline{(c)} \subset S_z^{(D)}$ $(z = e^{i\theta_\mu})$. In the opposite case, $z(e^{i\varphi})$ belongs to E for every $e^{i\varphi} \in E_\zeta$. As E is of linear measure zero, E_ζ must be of linear measure zero, by a well-known extension of Löwner's lemma.[7] In other words, $(w(\zeta) - w_0)/\rho$ is a function of class (U) in Seidel's sense.[8] Hence we have $\overline{(c)} \subset \overline{f(\Delta_\mu)}$. Since there is an infinite number of distinct components $\{\Delta_\mu\}$ converging to z_0, we have $\overline{(c)} \subset S_{z_0}^{(D)}$.

Next we consider a sequence of cross-cuts L_n $(n = 1, 2, \ldots)$ converging to z_0 and a corresponding sequence of positive numbers $\{\rho_n\}$ tending to zero, where L_n and ρ_n are chosen as above for each n. We denote by $\Delta_\mu^{(n)}$ the component, containing $z_\mu (\mu \geqq N(n))$, of the inverse image of (c_n): $|w - w_0| < \rho_n$. Suppose that there exists at least one n for which the sequence $\Delta_\mu^{(n)}$ $(\mu \geqq N(n))$ consists of infinitely many regions. We can conclude from the discussion above that $S_{z_0}^{(D)}$ contains $|w - w_0| \leqq \rho_n$. Thus it remains to consider the case in which, for every n, $\Delta_\mu^{(n)}$ consists of a finite number of different regions. Then it is easy to see that there exists an asymptotic path λ of $f(z)$ terminating at the point z_0 such that w_0 is the asymptotic value. In the case under consideration, we can show that if Ω denotes the complementary component of $\Gamma_{z_0}^{(C)}$ which contains w_0, then $R_{z_0}^{(D)}$ covers Ω, except for at most a set of capacity zero. Contrary to the assertion, suppose that $\Omega - R_{z_0}^{(D)}$ is of positive capacity. Then we can find a disk (K_0): $|z - z_0| < r_0$ and a set $\Sigma \subset \Omega$ of positive capacity, with the property that $f(z)$ does not belong to Σ for any $z \in D \cap (K_0)$. It is clear that there is a point $\alpha \in \Sigma$, different from w_0, such that for all positive numbers η the intersection $U(\alpha, \eta) \cap \Sigma$ is of positive capacity, $U(\alpha, \eta)$ denoting the η-neighborhood of α. Let α_1 be another point belonging to Σ and distinct from α and w_0. Inside Ω we draw a simple closed regular analytic curve Q

which surrounds w_0, α and passes through α_1 and whose interior G consists only of interior points of Ω. We put $\Sigma_1 = U(\alpha, \eta_1) \cap \Sigma$, choosing η_1 so small that $U(\alpha, \eta_1)$ is completely contained in G. Now, we select a positive number r ($< r_0$) such that the set M_r is outside the closure $\bar{G}$ of G. Denote by (K) the disk $|z - z_0| < r$. We construct a cross-cut L of $(K) \cap D$ which consists of two last parts $\Lambda'(e^{i\theta_1})$ and $\Lambda'(e^{i\theta_2})$ of $\Lambda(e^{i\theta_1})$ and $\Lambda(e^{i\theta_2})$ ($e^{i\theta_1}$, $e^{i\theta_2} \in \mathfrak{C}E$, $\theta_1 < \theta_0 < \theta_2$) and a rectilinear segment s, using the method stated above. We may suppose that the images of $\Lambda'(e^{i\theta_1})$ and $\Lambda'(e^{i\theta_2})$ have positive distances from $\bar{G}$. In the region D_1 bounded by L and the circular arc $(e^{i\theta}; \theta_1 \leqq \theta \leqq \theta_2)$, we consider the inverse image $f^{-1}(G)$ and denote by Δ the component of $f^{-1}(G)$ which contains the last part of the asymptotic path λ. The boundary of Δ consists of a finite number of segments q_i ($i = 1, 2, \ldots, m$)[9] on s and at most a countable number of analytic curves lying in D_1 and a bounded closed set on $|z| = 1$. Clearly Δ is simply connected. We map Δ by $z = z(\zeta)$ conformally upon the unit disk $|\zeta| < 1$ in a one-to-one manner. The image of λ terminates at a point ζ_0 on $|\zeta| = 1$. We consider a small open arc A_ζ of $|\zeta| = 1$ containing ζ_0 and disjoint from all σ_i, where σ_i denotes the image of q_i. Now, we consider the subset E_ζ of A_ζ such that, for $\zeta = e^{i\varphi} \in E_\zeta$, both $z = z(\zeta)$ and $w = w(\zeta) = f[z(\zeta)]$ have radial limits $z(e^{i\varphi})$ and $w(e^{i\varphi})$, and such that $w(e^{i\varphi}) \in G$. Suppose that there exists a point $e^{i\varphi} \in E_\zeta$ such that $z(e^{i\varphi})$ coincides with $z = e^{i\theta} \in \mathfrak{C}E$. We put $\beta = w(e^{i\varphi})$. Then, by a classical theorem on cluster sets, $\bar{G} \subset S_z^{(D)}$, and every value of $G - (\beta)$ is taken by $f(z)$ infinitely often in any neighborhood of $z = e^{i\theta}$, except for at most two values of $G - (\beta)$. This is a contradiction. Accordingly, E_ζ must be of linear measure zero, and for almost every $e^{i\varphi} \in A_\zeta$ the radial limit $w(e^{i\varphi})$ of $w(\zeta)$ lies on Q. Hence, if we denote by $W = \Phi(w)$ a function mapping G conformally upon the unit disk $|W| < 1$ in a one-to-one manner, the composed function $W = W(\zeta) = \Phi[f(z(\zeta))]$ is a function of generalized class (U). That is, the function $W = W(\zeta)$ is analytic and bounded: $|W(\zeta)| < 1$ in $|\zeta| < 1$, and the modulus of the radial limit $W(e^{i\varphi})$ is equal to 1 for almost every $e^{i\varphi}$ in the open arc A_ζ. Furthermore, $W = W(\zeta)$ has a radial limit $W(e^{i\varphi_0})$, $\zeta_0 = e^{i\varphi_0}$, which is equal to $\Phi(w_0)$ lying in $|W| < 1$. By an extension of Seidel-Frostman's theorem,[10] $W(\zeta)$ takes on every value of $|W| < 1$ infinitely often in any neighborhood of ζ_0, except for at most a set of values of capacity zero. This is a contradiction, since the image $\Phi(\Sigma_1)$ of Σ_1 is a set of exceptional values of positive capacity.

3. By arguments similar to those used in the proof of Theorem 1, we can prove

THEOREM 2. *If $\alpha \in S_{z_0}^{(D)} - \Gamma_{z_0}^{(C)}$ is an exceptional value of $f(z)$ in a neighborhood of z_0, then either α is an asymptotic value of $f(z)$ at z_0, or there exists a sequence of points $z_n' \in C$ converging to z_0 such that α is an asymptotic value of $f(z)$ at each z_n'.*

THEOREM 3. *If $\Omega_0 = S_{z_0}^{(D)} - \Gamma_{z_0}^{(C)}$ is not empty, then the range of values $R_{z_0}^{(D)}$ of $f(z)$ covers Ω_0, with a possible exceptional set of capacity zero.*

THEOREM 4. *Let H be a given closed set in the w-plane, Ω_1 a component of the complement $\mathfrak{C}H$ of H with respect to the w-plane, and α an accessible boundary point of Ω_1. Let the cluster set $S_\Lambda(e^{i\theta})$ be included in H for every $e^{i\theta} \in \mathfrak{C}E$, and let $S_{z_0}^{(D)} \cap \Omega_1 \neq \emptyset$. Furthermore, let the set Ξ of $e^{i\theta} \in \mathfrak{C}E$ with $\alpha \in S_\Lambda(e^{i\theta})$ be of linear measure zero. Then, if α is an exceptional value of $f(z)$ in a neighborhood of z_0, α is an asymptotic value of $f(z)$ arbitrarily near z_0.*

* This work was supported by a research project at Harvard University, sponsored by the Office of Naval Research, under contract No. N5ori-07634. The author wishes to express his hearty thanks to Professors Ahlfors and Lohwater for their kind suggestions.

[1] M. Ohtsuka, "On the Cluster Sets of Analytic Functions in a Jordan Domain," *J. Math. Soc. Japan*, **2**, 1–15, 1950; cf. Theorem 1.1.

[2] A. J. Lohwater, (1) "The Boundary Values of a Class of Meromorphic Functions," *Duke Math. J.*, **19**, 243–252, 1952; cf. Theorems 6 and 8; (2) "Les valeurs asymptotiques de quelques fonctions méromorphes dans le cercle-unité," *Compt. rend. Acad. sci.* (Paris), **237**, 16–18, 1953.

[3] O. Lehto, "On Meromorphic Functions Whose Values Lie in a Given Domain," *Ann. Acad. Aci. Fennicae*, Ser. A.I., **160**, 1953; cf. Theorem 5 and Corollary 2.

[4] D. A. Storvick, "On Meromorphic Functions of Bounded Characteristic," *Bull. Am. Math. Soc.*, **60**, 481, 1954.

[5] Cf. K. Noshiro, (1) "On the Theory of the Cluster Sets of Analytic Functions," *J. Fac. Sci., Hokkaido Univ.*, **6**, 217–231, 1938; (2) "Note on the Cluster Sets of Analytic Functions," *J. Math. Soc. Japan*, **1**, 275–281, 1950.

[6] K. Noshiro, *J. Math. Soc. Japan*, **1**, 275–281, 1950.

[7] Cf. S. Kametani and T. Ugaeri, "A Remark on Kawakami's Extension of Löwner's Lemma," *Proc. Acad. Tokyo*, **18**, 14–15, 1942.

[8] W. Seidel, "On the Distribution of Values of Bounded Analytic Functions," *Trans. Am. Math. Soc.*, **36**, 201–226, 1934; O. Frostman, *Potentiel d'équilibre et capacité des ensembles* (Lund, 1935), pp. 103–115.

[9] Some of them may degenerate to single points.

[10] O. Frostman, *op. cit.*

INTERACTIONS BETWEEN ELECTRONS AND WAVES

BY L. BRILLOUIN

COLUMBIA UNIVERSITY

Communicated March 21, 1955

1. *Introduction.*—A variety of problems relating to different technical devices have led to the discussion of the interaction between electrons and an electromagnetic wave. This corresponds to the cases of the linear accelerator or of the traveling wave amplifier, for instance.[1] These problems were discussed from the point of view of classical mechanics, and the most complete set of equations was obtained in the discussion of the traveling wave amplifier, where the reaction of the electron beam upon the wave is taken into account. The discussion of linear accelerators is usually based on the assumption of a very strong wave acting on a weak beam, and the reaction of the beam on the wave is neglected. The classical theory shows that electrons may be trapped by the wave and bunched. From there on, they travel with the wave velocity.[2] During the initial stage of trapping and bunching, the wave is split into different components, one of which is amplified while the others are attenuated. The final stage should yield a distorted wave of constant amplitude, with electrons bunched in the regions of low potential energy and traveling with the wave. Different types of stable solutions have been suggested, but the transition from the initial to the final stage is difficult to discuss, on account of the nonlinearity of the classical problem.

The situation is different when electronic motions are considered from the point of view of wave mechanics. We shall discuss this problem and show that it directly yields the stable solution of the final stage of interaction. The wave-mechanical problem appears as a generalization of the Mathieu equation, which corresponds to electrons moving in a fixed sinusoidal potential.

2. *Wave-mechanical Formulation.*—We consider the structure represented on Figure 1, which was discussed from a classical point of view earlier.[2] The notations are

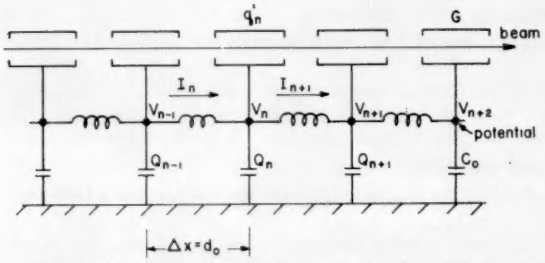


FIG. 1.—An electric line L_0C_0, coupled with an electronic beam, is used as a model for interaction between electrons and electromagnetic waves.

$L_0 = Ld_0$, self-induction per Section d_0,
$C_0 = Cd_0$, capacity per Section d_0,
I_n, current in L_0, Section n,
V_n, potential, Section n,
W_0, unperturbed wave velocity,
$LCW_0^2 = 1$. (1)

We may start from the set of equations previously established in the classical problem.[3]

$$\frac{\partial}{\partial x}(\rho v_x) + \frac{\partial \rho}{\partial t} = 0 \qquad \text{(conservation of electricity, } v_x \text{ electron velocity, } \rho \text{ electric density, } \rho v_x \text{ electric current density),} \tag{2}$$

$$\frac{dv_x}{dt} = -\frac{e}{m}\frac{\partial V}{\partial x} \qquad \text{(law of motion),}$$

$$\frac{\partial^2 V}{\partial x^2} - \frac{1}{W_0^2}\frac{\partial^2 V}{\partial t^2} = -L\frac{\partial^2 \rho}{\partial t^2} \qquad \text{(wave propagation with coupling to the electron beam).} \tag{3}$$

The conservation of electricity is automatically satisfied with usual quantum definitions of charge density and current. The law of motion (2) is replaced by a Schrödinger equation,

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar i}\frac{\partial \psi}{\partial t} - \frac{2m}{\hbar^2} eV(x, t)\,\psi = 0, \tag{4}$$

where V is the local potential, including the perturbation due to the electromagnetic wave. Equation (3) remains, with its right-hand term representing the coupling between the electromagnetic wave and electron. L is the self-induction of the line per unit length and plays the role of the coupling coefficient; ρ is $e\psi\psi^*$ in wave mechanics.

Let us first assume a weak coupling (L small). The wave propagates with a phase velocity W which may differ from the unperturbed W_0:

$$V = V_0 + 2U_0 \cos(\omega' t - k'x), \qquad \text{with } \omega' = k'W,\ W \approx W_0, \tag{5}$$

and equation (4) yields

$$\left[\frac{\partial^2}{\partial x^2} + \frac{2m}{\hbar i}\frac{\partial}{\partial t} + V_1 + U_1 e^{i\varphi'} + U_1 e^{-i\varphi'}\right]\psi(x, t) = 0, \tag{6}$$

with $\varphi' = \omega' t - k' x$, $V_1 = -(2me/\hbar^2)V_0$, $U_1 = -(2me/\hbar^2)U_0$. This is the basic equation which we must discuss.

This equation bears similarity to a Mathieu equation and reduces to the Mathieu type when

$$\omega' = 0, \qquad W = 0 \qquad \text{(Mathieu type)}, \quad (7)$$

A similar problem was stated by Parmenter,[4] who obtained an equation similar to (6) and discussed it with the method of Wannier and Slater.

3. *Discussion of the Interaction Equation.*—If we start with a ψ term in $e^{i\varphi_0}$, the U_0 terms in the equation immediately produce terms in $e^{i(\varphi_0 \pm \varphi')}$, and we end up with an infinite sum

$$\psi(x, t) = \sum_n A_n e^{i(\varphi_0 + n\varphi')} = \sum_n A_n e^{i\varphi_n},$$

$$\varphi_0 = \omega_0 t - k_0 x, \qquad \varphi_n = \omega_n t - k_n x, \quad (8)$$

$$\varphi' = \omega' t - k' x, \qquad \omega_n = \omega_0 + n\omega', \qquad k_n = k_0 + nk'$$

(A_n constant coefficients).

Factorizing $e^{i\varphi_0}$, we have

$$\psi(x, t) = e^{i\varphi_0} u_0(\varphi'),$$
$$u_0 = \Sigma A_n e^{in\varphi'}, \qquad \text{periodic in } \varphi' \text{ (period } 2\pi); \quad (9)$$

but we may just as well start from

$$\varphi_p = \varphi_0 + p\varphi'$$

and write

$$\psi(x, t) = e^{i\varphi_p} u_p(\varphi')$$
$$u_p = \Sigma A_n e^{i(n-p)\varphi'}, \qquad \text{periodic in } \varphi'. \quad (10)$$

Relations (9) and (10) are generalizations of the Floquet-Bloch theorem for Mathieu equations. The coefficients A_n in expansion (8) must satisfy an infinite set of simultaneous linear equations

$$\left[-k_n^2 + \frac{2m}{\hbar}\omega_n + V_1\right] A_n + U_1(A_{n+1} + A_{n-1}) = 0, \quad (11)$$

where we shall use the abbreviation

$$[n] = \left[-k_n^2 + \frac{2m}{\hbar}\omega_n + V_1\right]. \quad (11a)$$

These linear equations can be solved when the infinite determinant Δ is zero:

$$\Delta = \begin{vmatrix} \cdots & 0 & U_1 & [-1] & U_1 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & U_1 & [0] & U_1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & U_1 & [1] & U_1 & 0 \\ & & & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & \cdots & 0 & U_1 & [n] & U_1 & 0 \\ & & & & \cdots & \cdots & \cdots & \end{vmatrix} = 0. \quad (12)$$

This represents the generalization of Hill's determinant for the Mathieu problem. By solving this determinantal equation, we obtain a relation between ω_0 and k_0:

$$\Delta = F(\omega_0, k_0) = 0, \tag{13}$$

and for each k_0 value we find a series of solutions ω_0 (in fact, an infinite series). It is, however, obvious, from the similarity of equations (9) and (10) that the same solution could be obtained with ω_p and k_p; hence, when equation (13) is satisfied, we must also have

$$F(\omega_p, k_p) = F(\omega_0 + p\omega', k_0 + pk') = 0 \qquad (p \text{ integer}).^5 \tag{14}$$

This is the case if we take

$$\omega = Wk + f(k), \qquad f \text{ periodic in } k \text{ with period } k', \tag{15}$$

with $W = \omega'/k'$. The result is easy to check. Let ω_0 be a solution of equation (15) corresponding to k_0, and consider

$$\omega = \omega_0 + p\omega', \qquad k = k_0 + pk'. \tag{16}$$

The couple of values ω, k also satisfies equation (15), which proves that equations (13) and (14) are satisfied together.

Condition (15) reduces to a well-known result for the Mathieu equation, when $W = 0$.

4. *The Case of Small Coupling.*—We are specially interested in the case where the coupling between wave and electron is very small. This means that U_0 of equation (6) is a small quantity. Let us call A_0 the dominant term in expantion (8). The recurrence equations (11) indicate that A_n will be of the order of $U_0{}^n A_0$. We may work out an approximation keeping all terms up to $U_0{}^2$ and neglecting terms in $U_0{}^3$, $U_0{}^4$, ... This leaves us with five simultaneous linear equations (11), $n = -2, -1, 0, 1, 2$, to determine the unknown quantities A_{-2}, A_{-1}, A_1, A_2, and ω_0, when A_0, k_0 are given. The solution is readily obtained:

$$A_{\pm 1} = \frac{-U_1 A_0}{[\pm 1] - U_1{}^2/[\pm 2]}, \qquad A_{\pm 2} = \frac{U_1{}^2 A_0}{[\pm 1]\,[\pm 2] - U_1{}^2}, \tag{17}$$

and

$$[0] - \frac{U_1{}^2[2]}{[1]\,[2] - U_1{}^2} - \frac{U_1{}^2\,[-2]}{[-1]\,[-2] - U_1{}^2} = 0, \tag{18}$$

where the brackets have been defined in equation (11). The successive approximations are as follows:

Zero order:
$$[0] = 0, \tag{19}$$

First order:
$$[0] - U_1{}^2\left(\frac{1}{[1]} + \frac{1}{[-1]}\right) = 0, \tag{20}$$

Second order:

$$[0] - U_1{}^2\left(\frac{1}{[1]} + \frac{1}{[-1]}\right) - U_1{}^4\left(\frac{1}{[1]^2\,[2]} + \frac{1}{[-1]^2\,[-2]}\right) = 0, \tag{21}$$

where the U_1^4 terms result from expanding the fractions in equation (18) and may not be complete because of previous approximations.

Let us explain the meaning of these equations. Equation (19) represents the parabolic curve (see eqs. [8]–[11]) on Figure 2;

$$\omega_0 = \frac{\hbar}{2m}(k_0^2 - V_1) \tag{22}$$

for free electrons; and another condition $[n] = 0$ means

$$\omega_n = \frac{\hbar}{2m}(k_n^2 - V_1), \qquad \omega_n = \omega_0 + n\omega', \; k_n = k_0 + nk', \tag{23}$$

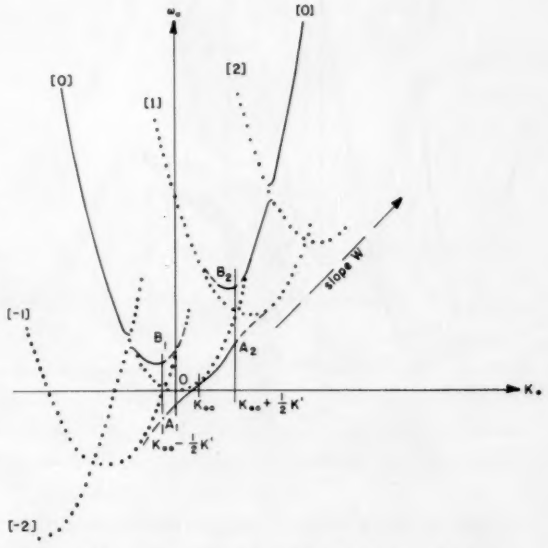


FIG. 2.—A case of weak coupling. The frequency ω_0 is proportional to the energy of the electron in the beam, while k_0 is proportional to the momentum of the electron. W is the velocity of the electromagnetic wave. The unperturbed [0] parabolic curve (dotted curve) is perturbed by the conpling with the wave and discontinuities appear at A_1B_1, A_2B_2 (solid curve). The median point k_{00} corresponds to electrons moving with a velocity equal to the wave velocity W.

Our zero-order approximation means free electrons (eq. [22]). The first approximation shows a correction, which becomes important only when

$$[1] \approx 0 \qquad \text{or} \qquad [-1] \approx 0,$$

and this happens at the crossings between the parabola [0] and the parabolas [1] or [−1].

The second-order correction is incomplete but indicates corrections of a higher order when either

$$[0] \approx 0, \; [2] \approx 0, \qquad \text{or } [0] \approx 0, \; [-2] \approx 0.$$

These results are shown on Figure 2. Intersections between the parabolas [−1], [0], [1] occur at $k_{00} \pm {}^1/_2k'$, with

$$k_{00} = \frac{mW}{\hbar}, \qquad v_{00} = \frac{\hbar k_{00}}{m} = W, \tag{24}$$

a condition which means electrons and waves traveling together with the same velocity.

Our approximation singled out a certain curve [0] as the point of departure. This is a slightly artificial procedure. All curves are exactly equivalent, and we may visualize this fact by repeating the curve periodically, with translations given by the vector OO' (components ω', k', slope W). This is shown on Figure 3, which illustrates the periodicity specified in equations (16).

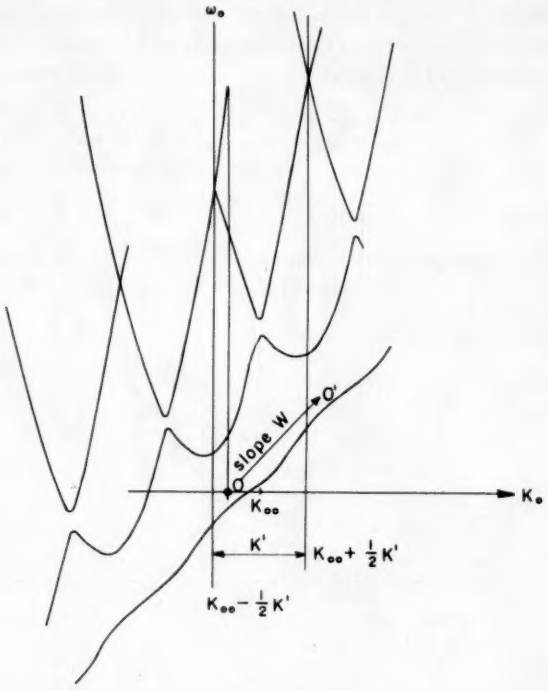


FIG. 3.—A different representation of the results of Fig. 2, using a k_0 value in the interval $k_{00} - {}^1/_2 k'$ to $k_{00} + {}^1/_2 k'$.

Since all these curves are just repetitions of each other, we may select definitions avoiding unnecessary redundancy. The definitions underlying the curve of Figure 1 were as follows:

$$k_0 \text{ varying from } -\infty \text{ to } +\infty, \tag{25}$$

$$\omega_0 \text{ discontinuous at } k_{00} + \frac{n}{2} k', \qquad n \text{ positive or negative integer.}$$

On Figure 3 we use another choice and assume k_0 to be restricted within a band of width k', for instance:

$$k_{00} - \frac{1}{2} k' \leq k_0 \leq k_{00} + \frac{1}{2} k', \tag{26}$$

$$\omega_0 \text{ discontinuous on the limits of the band.}$$

We may now consider the reaction on the electromagnetic wave, which is represented by the right-hand term in equation (3). Let us rewrite this equation for the case defined in equations (5) and (6):

$$\frac{\hbar^2}{me} U_1 \cos(\omega' t - k' x) \cdot \omega'^2 \left[\frac{1}{W^2} - \frac{1}{W_0^2}\right] + Le \frac{\partial^2}{\partial t^2} (\psi\psi^*) = 0, \tag{27}$$

since $k' = \omega'/W$ and $W \neq W_0$. We must now compute $\psi\psi^*$, keeping terms up to U_1 and dropping small terms in U_1^2 that would only mean nonlinear distortion. We thus keep terms 0, ± 1 only in equations (8) and (17):

$$\psi = A_0 e^{i(\omega_0 t - k_0 x)} + A_1 e^{i(\omega_1 t - k_1 x)} + A_{-1} e^{i(\omega_{-1} t - k_{-1} x)}$$
$$\omega_1 = \omega_0 + \omega', \qquad \omega_{-1} = \omega_0 - \omega', \tag{28}$$
$$k_1 = k_0 + k', \qquad k_{-1} = k_0 - k', \qquad A_{\pm 1} \approx \frac{-U_0 A_0}{[\pm 1]},$$

and we may assume A_0 real, hence $A_{\pm 1}$ real.

The ψ expansion being limited this way, we obtain

$$\psi\psi^* = A_0^2 + A_0 A_1^* e^{-i(\omega' t - k' x)} + A_0^* A_1 e^{i(\omega' t - k' x)} + A_0 A_{-1}^* e^{i(\omega' t - k' x)} + A_0^* A_{-1} e^{-i(\omega' t - k' x)} + \ldots, \tag{29}$$
$$\psi\psi^* = A_0^2 + C \cos(\omega' t - k' x),$$

with

$$C = 2A_0(A_1 + A_{-1}) = -2A_0^2 U_1 \left\{ \frac{1}{[+1]} + \frac{1}{[-1]} \right\}.$$

This term is linear in U_0 but quadratic in A_0. Equations (27) and (29) now yield

$$\frac{1}{W^2} = \frac{1}{W_0^2} - \frac{2Le^2 m A_0^2}{\hbar^2} \left\{ \frac{1}{[+1]} + \frac{1}{[-1]} \right\}, \tag{30}$$

a relation showing that the actual wave velocity W will differ slightly from the unperturbed W_0 on account of the interaction with electrons. The denominators $[\pm 1]$ are small near the discontinuities $k_{00} \pm \frac{1}{2}k'$ in the curve of Figure 2, indicating that these discontinuities result from a rather strong interaction between electron and wave.

5. *Physical Interpretation for the Case of Small Coupling.*—The physical meaning of the preceding results can be obtained by comparison with the classical problem:

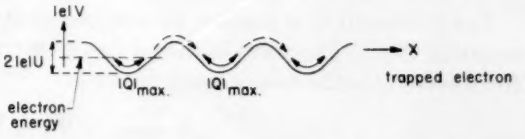


FIG. 4.—Electrons moving in the wave, case of trapped electrons.

A. The case of a *trapped electron* is sketched in Figure 4. The electron normally oscillates in a valley of potential energy. Occasionally this electron may jump from one valley to the next one by a "tunnel" effect. Altogether, the electric density $|\rho|$ is maximum in the valley and minimum on the hills. The average energy is in the $2|e|V_0$ band.

B. An *electron* with higher energy will *move freely* through the hills and valleys, moving faster in the valleys and slower on top of the hills. Here the maximum density $|\rho|$ will appear on the hills (Fig. 5).

This gives us a criterion for distinguishing between trapped and free electrons. According to equation (5), the potential energy of an electron is

$$eV = -|e|V = -|e|V_0 - 2|e|U_0 \cos\varphi', \qquad \varphi' = \omega' t - k' x,$$

or

$$\frac{2m}{\hbar^2} eV = -V_1 - 2U_1 \cos \varphi'. \tag{31}$$

Let us assume $U_0 > 0$, hence $U_1 > 0$; we note that the "valleys" correspond to

$$\varphi = 0, \pm 2\pi \ldots \pm 2n\pi \qquad \text{valleys of potential energy.} \tag{32}$$

The electronic density was computed in equations (29). The distinction between trapped or free electrons depends upon the sign of the coefficient C:

$$\begin{array}{lll} \text{A: Trapped electrons, } C > 0, & & \psi\psi^* \text{ max. in the valleys,} \\ \text{B: Free electrons, } C < 0, & & \psi\psi^* \text{ min. in the valleys.} \end{array} \tag{33}$$

The formula for C contains the two quantities $[+1]$ and $[-1]$ which were defined in equation (11a). The bracket $[+1]$, for instance, is positive above the parabola [1] of Figure 2 and negative below the parabola. The sign of C depends upon

FIG. 5.—Free electrons moving through the wave.

the position of the final curve with respect to both parabolas $[+1]$ and $[-1]$. The lower part, A_1A_2, of the final curve is comprised between $k_{00} - {}^1/_2k' \leq k_0 \leq k_{00} + {}^1/_2k'$ and lies below both parabolas. Hence $[+1]$ and $[-1]$ are negative, and C is positive along this part of the curve:

$$\begin{array}{lll} \text{A: Trapped electron, } C > 0, & & k_{00} - {}^1/_2k' \leq k_0 \leq k_{00} + {}^1/_2k', \\ \text{B: Free electron, } C < 0, & & \text{for other parts of the curve.} \end{array} \tag{34}$$

The sign of C on the other parts of the curve is determined by the sign of $\{(1/[+1]) + (1/[-1])\}$ and can be easily checked.

The coefficient C is positive for trapped electrons; hence $W < W_0$, according to equation (30). The wave is slowed down by trapped electrons and accelerated by free electrons, in the case of small coupling:

$$W_{\text{trapped}} < W_0 < W_{\text{free}}. \tag{35}$$

Let us investigate more carefully the situation at A_2B_2 when

$$k = k_{00} + {}^1/_2k'. \tag{36}$$

The brackets $[0]$ and $[+1]$ are small, while $[-1]$ is negative and large. Equation (20) practically reduces to

$$[0]\,[1] = U_0^2, \tag{37}$$

and at point (36), where both curves [0] and [1] intersect, the brackets are equal:

$$[0] = [1] = \pm U_0. \tag{38}$$

We may refer to definition (11a) of the abbreviations and call ω_0 the frequency at the intersection point

$$k = k_{00} + {}^1/_2k', \qquad \omega = \omega_0, \qquad [0] = [1] = 0;$$

the actual points A_2B_2 correspond to

$$\omega = \omega_0 \pm \epsilon, \qquad [0] = [1] = \pm \frac{2m}{\hbar} \epsilon;$$

hence

$$\epsilon = \frac{\hbar}{2m} U_1, \qquad \text{or} \qquad \hbar\epsilon = |e| U_0, \tag{39}$$

with the help of equation (6). This expression will be used later. From (38) and (30) we compute

$$\frac{1}{W^2} - \frac{1}{W_0^2} = \mp \frac{2Le^2m}{\hbar^2 U_1} A_0^2 = \mp \frac{L|e|}{U_0} A_0^2.$$

The velocity at B_2 (free electron) is larger than at A_2 (trapped electron), and our result means

$$W_{B_2} - W_{A_2} = \Delta W = \frac{2L|e|W_0^3}{U_0} A_0^2 = \frac{2L|e|W_0^3}{U_0 l}, \qquad \text{with } A_0^2 l = 1, \tag{40}$$

assuming one electron per length l of the structure in average. On the other hand, if we may assume a weak electron beam, the energy of the electromagnetic wave, per section d_0 of the line, can easily be computed from definitions (1):

$$\bar{E}_{d_0} = \frac{1}{2} L_0 \bar{I}^2 + \frac{\bar{Q}^2}{2C_0} = L_0 \bar{I}^2, \tag{41}$$

where I = current in the line and Q = charge on the capacity. Both terms in (41) are equal in average, and, furthermore (see eq. [5]),

$$L_0 \frac{\partial I}{\partial t} = d_0 \frac{\partial V}{\partial x} = 2d_0 U_0 k' \sin \varphi', \qquad \varphi' = \omega' t - k' x;$$

hence

$$I = -\frac{2V_0}{LW_0} \cos \varphi',$$

and we obtain the average energy per unit length of the line (neglecting charge in the electron beam):

$$\bar{E}_1 = \frac{1}{d_0} \bar{E}_{d_0} = L\bar{I}^2 = \frac{2U_0^2}{LW_0^2}. \tag{42}$$

Using this value in equation (40), we have

$$\frac{\Delta W}{W_0} = 4 \frac{U_0|e|}{\bar{E}_1 l}, \tag{43}$$

and this represents the change in wave velocity when a trapped electron becomes free ($A_2 \rightarrow B_2$ transition).

This result will enable us to compute the *total change in energy during the* $A_2 \rightarrow B_2$ *transition.* The change on electron energy from A_2 to B_2 results from equation (39) and amounts to

$$2\hbar\epsilon = 2|e|U_0; \tag{44}$$

but we have also to account for the change ΔW in the wave velocity. Boundary conditions will keep k' constant. If, for instance, we select a cyclic condition over a length l, as in equation (40), we shall have a condition

$$k' = 2\pi \frac{q}{l} \qquad (q \text{ integer}). \tag{45}$$

A change in velocity W will then mean a change in frequency

$$\frac{\Delta\omega'}{\omega'} = \frac{\Delta W}{W}. \qquad \text{(This will be the situation in solid-state problems.)} \tag{46}$$

In a technical problem, on the contrary, the frequency ω would be given and subject to no change.

Condition (45) results in a selection of energy levels

$$\bar{E}_2 l = (r + {}^1/_2)\hbar\omega' \qquad (r \text{ integer})$$

for the energy over the distance l; hence

$$\frac{\Delta(\bar{E}_2 l)}{\bar{E}_2 l} = \frac{\Delta\omega'}{\omega'} = \frac{\Delta W}{W} = \frac{4|e|U_0}{\bar{E}_2 l}, \tag{47}$$

according to equation (43).

This means that freeing an electron ($A_2 \rightarrow B_2$ transition) requires a total energy given by equations (44) and (47):

$$\Delta E = 2|e|U_0 + \Delta(\bar{E}_2 l) = 6|e|U_0. \tag{48}$$

The change in the wave velocity involves twice as much energy as the discontinuity in the electron's energy curve of Figure 2.

6. *Solution by Reduction to a Mathieu Equation.*—We discussed in Section 3 some general properties of our equation (6) and investigated in Sections 4 and 5 the special case of small coupling. A general solution can be obtained by reducing our problem to a Mathieu equation. This is done by the following change of function:

$$\psi(x, t) = e^{i(\omega t - k_{00}x)}\,\Phi(\xi), \tag{49}$$

with $\xi = \varphi'/_2 = {}^1/_2(\omega' t - k'x)$ and $k_{00} = mW/\hbar$. Using (49) in equation (6), we obtain for Φ the condition

$$\frac{\partial^2\Phi}{\partial\xi^2} + \Phi\,\frac{4}{k'^2}\left[\frac{2m}{\hbar}\,\omega - k_{00}^2 + V_1 + 2U_1\cos 2\xi\right] = 0, \tag{50}$$

to be compared with a Mathieu equation[6]

$$\frac{\partial^2\Phi}{\partial\xi^2} + \Phi(\eta + \gamma\cos 2\xi) = 0; \tag{51}$$

hence

$$\eta = \frac{4}{k'^2}\left[\frac{2m}{\hbar}\,\omega - k_{00}^2 + V_1\right], \qquad \gamma = \frac{8U_0}{k'^2}.$$

A solution of a Mathieu equation corresponding to a single sustained wave can be written

$$\Phi = A(\xi)e^{i\beta\xi} \qquad (\beta \text{ real}), \quad (52)$$

where $A(\xi)$ is a periodic function with period π in ξ. Reverting to our previous φ' variable,

$$\Phi = F(\varphi')e^{i(\beta/2)\varphi'} \qquad (F \text{ period } 2\pi \text{ in } \varphi'); \quad (53)$$

hence

$$\psi = e^{i[\omega+(\beta/2)\omega']t - i[k_{00}+(\beta/2)k']x}\, F(\omega' t - k'x) \quad (54)$$

for the general solution. Let us compare this with our previous expression (9). We obtain from this comparison

$$\omega_0 = \omega + \frac{\beta}{2}\,\omega', \qquad k_0 = k_{00} + \frac{\beta}{2}\,k'. \quad (55)$$

We may also replace the periodic function F by another function,

$$\begin{aligned} &F_p(\varphi') = F(\varphi')e^{ip\varphi'}, \\ &\Phi = F_p(\varphi')e^{i[(\beta/2)+p]\varphi'} \qquad (p \text{ integer}). \end{aligned} \quad (56)$$

A check with our previous equation (10) now gives

$$\begin{aligned} \omega_p &= \omega + \frac{\beta}{2}\,\omega' + p\omega' = \omega_0 + p\omega', \\ k_p &= k_{00} + \frac{\beta}{2}\,k' + pk' = k_0 + pk', \end{aligned} \quad (57)$$

a result which checks with the original one. We thus obtain a complete agreement with our general remarks of Section 3.

We may now draw some conclusions about the solutions for cases of strong coupling. The Mathieu equation yields a certain number of passing bands when β is real.[7] Two different notations can be used in this connection:

Passing band number		Continuous β variation
$0\ -1 < \beta < +1$	$\Delta\beta = 2$	$-1 < \beta < +1$
$1\ -1 < \beta < +1$		$-2 < \beta < -1$ and $1 < \beta < 2$
$9\ -1 < \beta < +1$		$-9 - 1 < \beta < -9$ and $9 < \beta < 9 + 1$

(58)

The first notation corresponds to the graph of Figure 3, while the notation with a continuously varying β corresponds to the case of Figure 2.

The first notation is more convenient for problems of strong coupling (γ large). The first few bands are very narrow and practically give for η (hence for ω) a constant value. Higher bands exhibit a dependence of ω upon β. The zero-th passing band, for instance, gives

$$\omega = \Omega_0, \qquad \omega_0 = \Omega_0 + \frac{\beta}{2}\omega', \qquad k_0 = k_{00} + \frac{\beta}{2}k', \qquad -\frac{1}{2} < \frac{\beta}{2} < +\frac{1}{2} \tag{59}$$

according to formulas (55) and (58). On the ω_0, k_0 diagram we obtain a straight line with a slope $W = (\omega'/k')$. The structure of a system of bands is indicated in Figure 6.

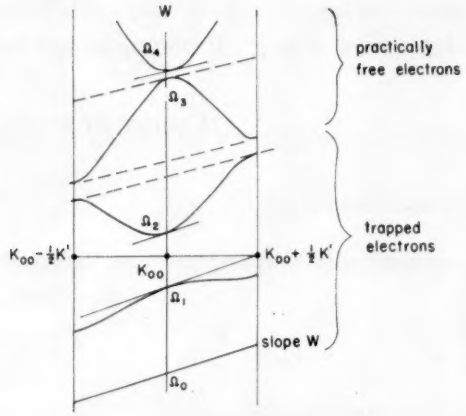


FIG. 6.—Case of strong coupling between electrons and waves. A succession of energy bands is easily recognized. The average slope of the bands is equal to the velocity W of the electromagnetic waves.

The distinction between trapped or free electrons was discussed for the Mathieu case and can be directly transcribed here:

I. $\eta < -\gamma$ (no passing band), $\dfrac{2m}{\hbar}\omega < k_{00}^2 - V_1 - 2U_1$,

II. $-\gamma < \eta < +\gamma$ (narrow passing bands [trapped electrons]),

$$k_{00}^2 - V_1 - 2U_1 < \frac{2m}{\hbar}\omega < k_{00}^2 - V_1 + 2U_1, \tag{60}$$

III. $\gamma < \eta$ (broad passing bands [free electrons]),

$$k_{00}^2 - V_1 + 2U_1 < \frac{2m}{\hbar}\omega.$$

These results give the basis for a discussion of strong coupling problems and agree with our previous investigation of the small coupling case.

The reaction on the electromagnetic wave (eq. [3]) would be much more complicated to discuss for the case of strong coupling, since $\psi\psi^*$ would not be a cos φ' function as in the case of small coupling (eq. [29]). This means that nonlinear distortion would play a dominant role. Altogether, the problem would require very careful discussion: we have built up a ψ solution for a regular sine wave of given velocity acting on an electron, but the reaction of the electron on the wave will distort the wave and introduce harmonics; hence the whole problem is not completely solved.

7. *Discussion.*—A complete solution of the problem, in the case of strong interaction, would require a discussion of *self-consistent* field. This could be done along the following line: First, one should assume an electromagnetic wave con-

taining an arbitrary number of upper harmonics. Instead of the simple sine wave assumed in equation (5), it would be necessary to use a complete Fourier series:

$$V = V_0 + \sum_p V_p e^{ip\varphi'} \quad (p \text{ integer}),\ \varphi' = \omega' t - k' x, \tag{61}$$

The ψ function would still be represented by an expansion (8), but the linear relations (11) would contain higher terms, and Hill's determinant Δ in equation (12) would have all sorts of nondiagonal terms. Solving this set of equations would yield the A_n coefficients of the ψ function, from which the $\psi\psi^*$ electronic density could be computed:

$$\psi\psi^* = \sum_p B_p e^{ip\varphi'}, \qquad B_p = \sum_n A_n A^*_{n-p}. \tag{62}$$

This expression should be introduced in equation (3) and would not modify the wave form, provided that the B_p coefficient would be proportional to the corresponding V_p term:

$$B_p = \alpha V_p \qquad (\alpha \text{ constant}). \tag{63}$$

Equation (3) would then give the corrected value for the wave velocity W. The self-consistency of the method is based on conditions (63).

We can also write directly a nonlinear equation for the whole problem. We assume a wave propagating with a given shape and a constant velocity W:

$$V = V(x - Wt), \tag{64}$$

an assumption similar to equation (61). Equation (3) reads

$$\frac{\partial^2 V}{\partial t^2}\left(\frac{1}{W^2} - \frac{1}{W_0^2}\right) = -L\frac{\partial^2 \rho}{\partial t^2} = L|e|\frac{\partial^2}{\partial t^2}(\psi\psi^*). \tag{65}$$

Integrating without adding any additional terms that would be forbidden by our assumption of a stable wave system, we have

$$V = V_0 + \frac{W_0^2 W^2}{W_0^2 - W^2} L|e|\psi\psi^*, \tag{66}$$

an expression that can be used in the ψ equation (4) and results in a nonlinear ψ equation.

8. *Applications.*—A number of important applications of the present theory can be found, even for the case of small coupling only. We discussed in the Introduction some technical problems leading to our theoretical discussion. The usual treatment led to serious difficulties and did not give a definite answer on the final shape of the wave and on the electron distribution. If no damping term is introduced in the electronic motions, the final state should result directly from the initial conditions. But such a simplification is unrealistic. Bound electrons coupled with the wave oscillate within a potential valley. They must radiate, and their oscillations would be damped. Classically, they should actually accumulate at the bottom of the potential valleys and yield a very large (even infinite) space charge density. In the wave-mechanical problem this difficulty is avoided. The oscilla-

tions within the valley cannot die out completely, and energy levels with $^1/_2$, $^3/_2$, . . . , oscillation quanta would represent the final stage. The wave-mechanical method is thus needed for a correct discussion of stable waves and electron densities corresponding to the final state of coupling, after the interaction between wave and electron has been proceeding for a long time.

In any practical application, it is very important to discuss whether the discontinuities (A_1B_1 and A_2B_2 in Fig. 2) in the energy curve are actually observable or not. This depends directly upon the time θ available for observation. The wave or the electron beam may have a finite mean free path and hence a finite mean lifetime. Let θ be this limit. The uncertainty principle states that the energy E can be observed only to within an error ΔE_u:

$$\Delta t = \theta, \qquad \Delta E_u \cdot \Delta t \approx h, \qquad \Delta E_u \approx \frac{h}{\theta}. \tag{67}$$

The curves drawn in Figures 2, 3, or 6 should now be considered as having a thickness ΔE_u. Discontinuities ΔE in the curve can be observed only if they are larger than the uncertainty ΔE_u:

$$\Delta E > \Delta E_u = \frac{h}{\theta}. \tag{68}$$

This is the condition defining when the peculiar structure of the curves will be observed. According to the discussion of Section 5 (small coupling), we have

$$\Delta E = 2 \cdot |e| \cdot U_0 \qquad \text{(eq. [44])}, \tag{69}$$

if we consider the frequency ω' of the wave as fixed. On the other hand, we found

$$\Delta E = 6 \cdot |e| \cdot U_0 \qquad \text{(eq. [48])} \tag{70}$$

for problems where the k' of the wave was given and the frequency ω' resulted from the variable wave velocity W.

In any case we summarize these results:

$$\Delta E = \alpha \cdot |e| \cdot U_0, \qquad \alpha = \begin{cases} 2 \\ 6, \end{cases} \tag{71}$$

with a coefficient α equal to 2 or 6, as the case may be. The special structure of the energy curve can be observed only when the duration θ of an observation is large enough so that

$$\theta > h/[\alpha \cdot |e| \cdot U_0] \tag{72}$$

according to equation (68) and (71). A similar result can be obtained directly. In order for the ΔE discontinuity to be observable, we must be able to observe "beats" between the two frequencies corresponding to an interval ΔE (points A_1B_1 or A_2B_2 in Fig. 2). These beats occur at intervals

$$\Delta t = \frac{2\pi}{\Delta\omega} = \frac{h}{\Delta E}, \tag{73}$$

since $\hbar\,\Delta\omega = \Delta E$. Observation is possible only when Δt is smaller than the time limit θ, a condition which leads directly to inequality (68).

Let us mention some other related subjects. Instead of electromagnetic waves, we may consider elastic waves in a conductor. These waves result in motions of the positive ions with alternating layers of increased or decreased density. This, in turn, creates an oscillating potential field traveling with the wave, and this field may interact with free electrons in the solid. Such a problem was stated by Parmenter,[4] and our discussion would apply directly to this case. Interaction between elastic waves and free electrons in a metal was also suggested by Bardeen[8] and Fröhlich[9] as a possible explanation of superconductivity. It is hoped that the present discussion will help in clarifying this difficult problem. According to condition (72), the elastic waves able to play a dominant role in the mechanism of superconductivity should be those having a very long mean free path and a long lifetime θ. In this respect, special attention should be given to the waves reduced to their zero-point energy of one-half quantum. These waves have abnormally large θ values and should play an important role in trapping free electrons and making them insensitive to thermal agitation.

[1] For a general discussion of these problems see L. Brillouin, *Phys. Rev.*, **74**, 90, 1948.

[2] L. Brillouin, *J. Appl. Phys.*, **20**, 1196, 1949.

[3] *Ibid.*, eqs. (4), (6), and (13).

[4] R. H. Parmenter, *Phys. Rev.*, **89**, 990, 1953.

[5] The result can be seen directly on the determinant Δ, since the change from $\omega_0 k_0$ to $\omega_p k_p$ simply means shifting all rows up by p units, and this does not change the value of the infinite determinant.

[6] L. Brillouin, *Wave Propagation in Periodic Structures* (New York: Dover Press, 1953), p. 172.

[7] *Ibid.*, pp. 173–177.

[8] J. Bardeen, *Phys. Rev.*, **79**, 167, 1950, and **80**, 567, 1951; *Rev. Mod. Phys.*, **23**, 268, 1951.

[9] H. Fröhlich, *Phys. Rev.*, **79**, 845, 1950; *Proc. Roy. Soc. (London), A*, **215**, 291, 1952.

NOTE ON THE SIMPLE COLLISION THEORY OF BIMOLECULAR REACTIONS

BY R. D. PRESENT

UNIVERSITY OF TENNESSEE

Communicated by George Scatchard, April 1, 1955

The simple collision theory treats bimolecular reactions by gas-kinetic methods, omitting internal degrees of freedom of the reacting molecules as well as steric effects. Two forms of the simple theory have been proposed, based on different assumptions and leading to different results.[1] In the first theory it is assumed that reaction occurs when the kinetic energy of relative motion E exceeds a specified value ϵ^*; the reaction probability is taken to be one above, and zero below, the threshhold energy. The reaction cross-section is discontinuous, since it vanishes below the threshold and is constant and equal to the gas-kinetic cross-section above it. The resulting formula for the number of reactive collisions per unit volume per unit time is

$$N = Ze^{-\epsilon^*/kT}\left(\frac{\epsilon^*}{kT} + 1\right), \tag{1}$$

where Z is the total number of collisions per unit volume per unit time and ϵ^* is the threshold or activation energy.

In the second theory a rigid-sphere model is used, and it is assumed that reaction occurs if the translational energy associated with the component of the relative velocity along the line of centers, i.e., the radial kinetic energy, at the instant before impact exceeds a specified value. Denoting this "line-of-centers energy at contact" by ϵ, the reaction probability is taken to be unity for $\epsilon > \epsilon^*$ and zero for $\epsilon < \epsilon^*$. The resulting formula is

$$N = Ze^{-\epsilon^*/kT}. \tag{2}$$

Since $\epsilon^* \gg kT$ in practice, e.g., $\epsilon^*/kT \sim 30$ for measurements of the hydrogen iodide decomposition, formulas (1) and (2) differ by a significant factor in their predictions of absolute reaction rates. It is found in many cases that formula (2) gives remarkably good agreement with experiment, i.e., the observed reaction rates determine collision diameters which are of reasonable orders of magnitude. Furthermore, the transition-state theory, when applied to the same problem, predicts an absolute reaction rate in agreement with equation (2).[2] Because of the rather artificial nature of the second collision theory, it appears of interest to derive equation (2) by kinetic-theory methods in a somewhat more general fashion, without introducing the "line-of-centers energy at contact." In the derivation which follows, explicit account is taken of the potential energy of interaction and its relation to the activation energy. Also, the probability of reaction or the reaction cross-section is obtained explicitly as a function of the total relative kinetic energy of translation.

We assume a very simple classical model for the formation of the activated complex. When the centers of the collision partners approach to within a distance $r = D$, it is assumed that the molecules "stick together" to form the complex. If there were no forces between the colliding molecules, the cross-section σ for the formation of the activated complex would be simply πD^2. However, at close distances of approach, the "overlap" forces produce a strong repulsion which is responsible for the activation barrier. Let $V(r)$ denote the spherically symmetric potential energy of mutual repulsion of the colliding molecules with $V(\infty) = 0$. Then $V(D)$ will represent the activation energy ϵ^* for this simple model. The distance between centers at closest approach is determined by the relative velocity v and the impact parameter b. Those and only those values of b for which the distance of closest approach would be less than D lead to formation of the activated complex. A simple calculation, *which is independent of the form of* $V(r)$, gives

$$\sigma(E) = \begin{cases} \pi D^2[1 - (\epsilon^*/E)], & E > \epsilon^*, \\ 0, & E < \epsilon^*, \end{cases} \tag{3}$$

where $E = \mu v^2/2$ is the complete translational energy of the relative motion, and μ is the reduced mass. Equation (3) predicts a hyperbolic rise of the cross-section beyond the threshhold with an asymptotic approach to the value πD^2. This is to be contrasted with the step-function

$$\sigma = \begin{cases} \pi D^2, & E > \epsilon^*, \\ 0, & E < \epsilon^*, \end{cases} \tag{4}$$

which is assumed in the first collision theory. Since the activated complex may in some cases decompose into the initial reactants rather than into the products of the reaction, the cross-section for reaction must be smaller than the activation cross-section. However, since the two cross-sections will usually differ by a factor of the order of unity, this factor will be considered to be absorbed in the already uncertain collision diameter D, and σ will be referred to as the reaction cross-section.

Using the well-known formula for the number of collisions per unit volume per unit time with the relative translational energy in the interval E to $E + dE$, one obtains for the number of reactive collisions per cubic centimeter per second in the same energy interval

$$dN = \frac{2\nu_1\nu_2}{(kT)^{3/2}} \cdot \left(\frac{2}{\pi\mu}\right)^{1/2} \cdot \sigma(E)e^{-E/kT}\, E\, dE, \tag{5}$$

where ν_1 and ν_2 denote number densities of molecules of different species. On inserting $\sigma(E)$ from equation (3) and integrating from ϵ^* to ∞, equation (5) becomes

$$\begin{aligned} N &= 2\nu_1\nu_2 D^2 \cdot \left(\frac{2\pi kT}{\mu}\right)^{1/2} \cdot e^{-\epsilon^*/kT} \\ &= Ze^{-\epsilon^*/kT}. \end{aligned} \tag{6}$$

Equations (6) and (2) are identical in form. The second collision theory also gives the same formula for dN as equation (5) with $\sigma(E)$ inserted from equation (3). The results are the same because the second theory is a special case of the foregoing in which the potential energy $V(r)$ is chosen to have the form of a step-function, the step of height $V(D)$ occurring at $r = D$. The condition that the distance of closest approach be less than D is equivalent to the condition that the radial kinetic energy at $r = D$ (i.e., the line-of-centers energy immediately before contact) be greater than $V(D)$.

It is believed that the model proposed here demonstrates the implausibility of the first collision theory and provides a more satisfactory basis for the results of the second collision theory. I am obliged to Dr. O. K. Rice for valuable comment on the manuscript.

[1] R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge: At the University Press, 1949), pp. 505–508.

[2] Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (New York: McGraw-Hill Book Co., 1941), pp. 16–18.

A THEORY CONCERNING THE MECHANISM OF ALLERGIC DISEASES

By Richard W. Lippman

RESEARCH DEPARTMENT, TEMPLE HOSPITAL, LOS ANGELES, CALIFORNIA*

Communicated by Linus Pauling, April 6, 1955

The causes of human glomerular nephritis and nephrosis usually have been considered to involve an allergic mechanism. Both diseases occur more commonly in those who have other manifestations of allergy, often members of families with an allergic diathesis. Acute nephritis usually follows infection with certain strains of hemolytic streptococcus,[1] with a latent period like that which would be necessary for antibody formation. During the nephrotic stage of glomerular nephritis the serum complement level falls precipitously, then rises as the nephrotic activity subsides.[2] Some have attributed nephritis to the immunologic effects of a hypothetical autogenous antibody to kidney tissue,[3] but the demonstration of autoantibodies to kidney tissue has been elusive.[4] There has always been a stumbling block in this explanation: Why should an autogenous tissue ever become antigenic? The association of streptococcal infection with nephritis has been invoked to overcome this stumbling block, and the antigenicity of autogenous kidney protein has been attributed to some interaction with streptococcal protein.[2, 5] Experiments to find out whether such an interaction can create autogenous or even homologous antigenicity have led to doubtful or negative conclusions.[5, 6]

Two types of experimental immunologic procedure will produce nephritis in animals. In one type, usually called "Masugi-type nephritis" (although the work of Masugi[7] was antedated by that of Lindemann[8] as well as by that of Wilson and Oliver,[9]), antikidney antibodies are formed in a heterologous species and then administered to produce nephritis in experimental animals. Obviously, this mechanism cannot be the natural one involved in the production of human nephritis. The second type of experimental nephritis is produced by administration of various foreign proteins, not themselves antibodies or extracted from kidney tissues.[10, 11, 12] Hawn and Janeway speculated provocatively concerning the reason why such foreign proteins, not themselves directly related to or derived from the kidney, should have the capacity to induce nephritis.[10]

In 1940 Kay observed that experimental nephritis in the rabbit followed a latent period after the administration of duck anti-rabbit-kidney serum and that the development of nephritis came simultaneously with appearance in the rabbit serum of antibodies to duck serum.[13] Indeed, he suggested that the duck anti-rabbit-kidney antibody combines with rabbit kidney tissue to produce a harmless combination (DAK-RK) which is fixed in the kidney tissue and that the rabbit anti-duck-serum antibodies, when formed, then combine with the harmless DAK-RK combination to produce a harmful antigen-antibody complex, which causes nephritis. Both types of experimental nephritis may be inhibited or enhanced by general factors affecting the response to any antigen-antibody reaction (state of adrenal function,[14] complement availability,[15] etc.). Kay's experiment may be considered as a special case of the second type of experimental nephritis, since the specific antikidney antibody produced no immediate effect, possibly as the result of a low antibody titer, and the subsequent nephritis seemed to result from the foreign character of duck serum protein.

By a slight extension, expressed in the following three hypotheses, the suggestions of Kay can be applied to explain the second type of experimental nephritis in a general way and can be applied, further, to human nephritis and even to other immunologic phenomena and diseases that follow the introduction of foreign antigens into the body.

FIRST HYPOTHESIS: *When a foreign antigen is introduced into the body, it may accumulate at certain local sites of predilection which are determined by the nature of the antigen, the route of introduction, and the amount of antigen.* It is evident that the sites of accumulation will be determined in part by the nature of the antigen. For example, a particulate antigen (such as a suspension of killed bacteria) might be expected to accumulate in the reticulo-endothelial system, while a soluble antigen might be expected to accumulate in the liver,[15] lymph tissue, or kidney. The route of introduction will determine the order in which the antigen reaches potential sites of accumulation. For example, an antigen that gains entrance to the peripheral circulation will travel to the lung through the pulmonary circulation; an antigen that is absorbed through the intestinal tract will travel mainly to the liver and, perhaps, partly through the lymph system and the systemic veins to the lung. The amount of antigen will determine in part the sites of accumulation, since a very small amount of antigen, regardless of its nature, might be fixed entirely in the first site reached; a very large amount of antigen might saturate the site of greatest affinity and accumulate thereafter in sites of secondary affinity.

SECOND HYPOTHESIS: *The antigen persists at the sites of predilection during the period when antibodies are formed.* Persistence of antigen for several weeks has been demonstrated by McMaster *et al.* in the rabbit and in the mouse.[16] Garvey and Campbell have demonstrated persistence of antigen for 8 weeks in the rabbit;[17] the persistence may be the result of an inability to utilize or destroy the antigen or may be the result of a slow rate of destruction. Following a single introduction, the amount of antigen at the sites may diminish; but the amount of antigen at these sites may be maintained or even increased through continuous or repeated introduction. The antibodies may be formed at the sites of antigen accumulation or in remote locations.

THIRD HYPOTHESIS: *After the antibodies are formed and released into the circulation, they react with the antigen, which is in greatest concentration at the sites of predilection, and thus produce a localized anaphylactoid (reversed Arthus) phenomenon, with inflammation, exudation, and other consequent results.*

The work of Smetana,[18] Oliver,[19] Rather,[20] and others has shown that the kidney is a site of predilection for the deposition of some proteins, among which are those which have been used to produce various forms of experimental nephritis. For example, Oliver, Moses, MacDowell, and Lee[21] demonstrated protein droplets in the tubular epithelium of rat kidney after the intraperitoneal administration of horse serum, and Wissler, Smull, and Lesh[12] found that glomerulitis, along with some other cardiovascular lesions, could be produced in the rabbit by a single intravenous administration of horse serum. In the latter experiment, and in other similar experiments concerning the second type of experimental nephritis,[10, 11] only pathologic criteria for glomerulitis were used, applied to the specimens obtained at autopsy when the experiment was concluded. However, Waugh and More[22] observed also the time when proteinuria appeared, after rabbits were given bovine serum gamma

globulin. They found a latent period of 8 days, consistent with the time required for the formation of antibodies to the foreign antigen. All these observations agree with the above theory.

If certain other foreign proteins, such as one produced by the proper strain of streptococcus, were also found to have a predilection for deposition in the kidney, the theory would then explain the pathogenesis of human nephritis as well. In view of the suspected relationship between streptococcal infection and such human diseases as rheumatic fever and rheumatoid arthritis, in addition to nephritis, it would seem profitable to investigate the deposition and disposal of streptococcal antigens as a next step.

* Professors Linus C. Pauling and Dan H. Campbell, of the California Institute of Technology, Pasadena, have contributed many criticisms and discussions of these ideas. Recent work which led to these speculations has been supported in part by a grant from the Johnson Fund of the American Philosophical Society.

[1] W. T. Longcope, *Bull. Johns Hopkins Hosp.*, **45**, 335, 1929; C. H. Rammelkamp, Jr., and R. S. Weaver, *J. Clin. Invest.*, **32**, 345, 1953.

[2] K. Lange, F. Graig, J. Oberman, L. Slobody, G. Ogur, and F. LoCasto, *Arch. Internal Med.*, **88**, 433, 1951.

[3] F. F. Schwentker and F. C. Comploier, *J. Exptl. Med.*, **70**, 223, 1939.

[4] C. F. Kay, P. F. Lucchesi, and R. B. Rutherford, *J. Immunol.*, **42**, 369, 1941.

[5] P. Cavelti and E. S. Cavelti, *Arch. Pathol.*, **40**, 158, 1945.

[6] E. Middleton, Jr., E. B. Middleton, B. C. Seegal, *A.M.A. Arch. Pathol.*, **56**, 125, 1953; H. Goodman, personal communication.

[7] M. Masugi and Y. Tomizuka, *Trans. Japan. Pathol. Soc.*, **21**, 329, 1931.

[8] W. Lindemann, *Ann. Inst. Pasteur*, **14**, 49, 1900.

[9] G. W. Wilson and J. Oliver, *J. Exptl. Med.*, **32**, 183, 1920.

[10] C. van Z. Hawn, and C. A. Janeway, *J. Exptl. Med.*, **85**, 571, 1947.

[11] R. H. More and D. Waugh, *J. Exptl. Med.*, **89**, 541, 1949; R. H. More and S. D. Kobernick, *A.M.A. Arch. Pathol.*, **51**, 361, 1951.

[12] R. W. Wissler, K. Smull, and J. B. Lesh, *J. Exptl. Med.*, **90**, 577, 1949.

[13] C. F. Kay, *J. Exptl. Med.*, **72**, 559, 1940.

[14] R. W. Lippman, H. U. Marti, E. E. Jacobs, and D. H. Campbell, *A.M.A. Arch. Pathol.*, **57**, 405, 1954.

[15] J. S. Garvey and D. H. Campbell, *J. Immunol.*, **72**, 131, 1954.

[16] P. D. McMaster and H. Kruse, *J. Exptl. Med.*, **94**, 323, 1951; P. D. McMaster, H. Kruse, E. Sturm, and J. L. Edwards, *J. Exptl. Med.*, **100**, 341, 1954.

[17] J. S. Garvey and D. H. Campbell, *Federation Proc.*, **14**, 463, 1955.

[18] H. Smetana, *Am. J. Pathol.*, **18**, 1029, 1942.

[19] J. Oliver, M. MacDowell, and Y. C. Lee, *J. Exptl. Med.*, **99**, 589, 1954.

[20] L. J. Rather, *J. Exptl. Med.*, **87**, 163, 1948.

[21] J. Oliver, M. J. Moses, M. MacDowell, and Y. C. Lee, *J. Exptl. Med.*, **99**, 605, 1954.

[22] D. Waugh and R. H. More, *J. Exptl. Med.*, **95**, 555, 1952.

ERRATA: ON THE MODULI OF RIEMANN SURFACES

In the article of the above title appearing in these PROCEEDINGS, **41,** 236–238, 1955, the following corrections should be made:

On page 236 in the last sentence $2p - 2$ should be replaced by p.

The second sentence of the last paragraph on page 237 should read " . . . not only S but also D . . . . "

In the same paragraph (p. 238) C_1 should read $\tilde{C}_1$.

The last sentence of this same paragraph (p. 238) should read "T is the reflection in L followed by the exchange of the two copies of D (the reflection in the real axis, rotation by π, translation back); and . . . "

H. E. RAUCH

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois. Articles of which a member of the Academy is author or co-author may extend to 10 printed pages, those by non-members to 3 printed pages; space above these limits will be billed to the author at a rate based on the actual composition costs of the article. Accompanying the script should be a statement by the author or his representative that he assumes responsibility for the charges if incurred.

GALLEY PROOF will be sent, with the understanding that charges for author's corrections shall be billed to him. Author should therefore make final revisions on the typewritten manuscripts. Page proofs will not be sent.

REPRINTS should be ordered when proof is returned, on the reprint order form which will be inclosed with the proof. The order blank gives a tabular estimation of the cost of reprints; a copy will be supplied at any time upon request.

CONTENTS

Page

BIOCHEMISTRY.—ZINC, A COMPONENT OF YEAST ALCOHOL DEHYDROGENASE . . . . . *By Bert L. Vallee and Frederic L. Hoch* 327

BIOCHEMISTRY.—PRODUCTION OF ABNORMAL (PSYCHOTIC?) BEHAVIOR IN MICE WITH LYSERGIC ACID DIETHYLAMIDE, AND ITS PARTIAL PREVENTION WITH CHOLINERGIC DRUGS AND SEROTININ . . . . . *By D. W. Woolley* 338

GENETICS.—FINE STRUCTURE OF A GENETIC REGION IN BACTERIOPHAGE . . . . . *By Seymour Benzer* 344

GENETICS.—CHROMOSOMAL ABERRATIONS IN *Datura* DUE TO VARIOUS KINDS OF IRRADIATION . . . . . *By Jean M. Cummings, Lewis Goldstein, and A. F. Blakeslee* 355

GENETICS.—EVIDENCE OF COMPLEX LOCI IN *Salmonella* . . . . . *By M. Demerec, I. Blomstrand, and Z. E. Demerec* 359

GENETICS.—GENETICALLY CONDITIONED METABOLIC ANTAGONISM AND SUPPRESSOR ACTION IN *Neurospora* . . . . . *By C. O. Doudney and R. P. Wagner* 364

GENETICS.—NUCLEO-CYTOPLASMIC RELATIONS IN SALIVARY-GLAND CELLS OF *Drosophila* . . . . . *By Helen Gay* 370

GENETICS.—PHENOTYPIC VARIATION AND PSEUDO-ALLELISM AT THE FORKED LOCUS IN *Drosophila Melanogaster*. . . . . *By M. M. Green* 375

MATHEMATICS.—CURVILINEAR CLUSTER SETS OF ARBITRARY FUNCTIONS . *By F. Bagemihl* 379

MATHEMATICS.—ON A CHARACTERIZATION OF THE NORMAL LAW . . . *By Glen Baxter* 383

MATHEMATICS.—ON THE ADDITIVITY OF THE TRACE IN FINITE FACTORS . . . . . *By Richard V. Kadison* 385

MATHEMATICS.—REPRESENTATION OF A CLASS OF STOCHASTIC PROCESSES . . . . . *By Samuel Karlin and James McGregor* 387

MATHEMATICS.—A VARIATIONAL METHOD IN THE THEORY OF HARMONIC INTEGRALS . . . . . *By Charles B. Morrey, Jr., and James Eells, Jr.* 391

MATHEMATICS.—ON THE CHARACTERS OF THE SYMMETRIC GROUP . . . . . *By Francis D. Murnaghan* 396

MATHEMATICS.—CLUSTER SETS OF FUNCTIONS MEROMORPHIC IN THE UNIT CIRCLE . . . . . *By Kiyoshi Noshiro* 398

PHYSICS.—INTERACTIONS BETWEEN ELECTRONS AND WAVES . . . . *By L. Brillouin* 401

PHYSICS.—NOTE ON THE SIMPLE COLLISION THEORY OF BIMOLECULAR REACTIONS . . . . . *By R. D. Present* 415

PHYSIOLOGY.—A THEORY CONCERNING THE MECHANISM OF ALLERGIC DISEASES . . . . . *By Richard W. Lippman* 418

ERRATA.—ON THE MODULI OF RIEMANN SURFACES . . . . . *By H. E. Rauch* 421